Genomic signatures of desert adaptation at gene-rich regions in zebu cattle from the African drylands

Abdulfatai Tijjani^{1,2,3*}, Bashir Salim⁴, Marcos Vinicius Barbosa da Silva⁵, Hamza A. Eltahir⁶,
Taha H. Musa⁷, Karen Marshall^{8,9}, Olivier Hanotte^{1,2,3*}, Hassan H. Musa^{6,10*}

¹ International Livestock Research Institute (ILRI) PO 5689, Addis Ababa, Ethiopia

² Centre for Tropical Livestock Genetics and Health (CTLGH), ILRI Ethiopia, PO Box 5689, Addis Ababa, Ethiopia

³ Cells, Organisms and Molecular Genetics, School of Life Sciences, University of Nottingham, United Kingdom

⁴ Faculty of Veterinary Medicine, University of Khartoum, Sudan

⁵ Embrapa Gado de Leite, Juiz de Fora, Brazil

⁶ Institute of Molecular Biology, University of Nyala, Sudan

⁷ Biomedical Research Institute, Darfur College, Sudan

⁸ International Livestock Research Institute (ILRI) PO Box 30709, Nairobi 00100, Kenya.

⁹ Centre for Tropical Livestock Genetics and Health (CTLGH), ILRI Kenya, P.O. Box 30709, Nairobi 00100, Kenya.

¹⁰ Faculty of Medical Laboratory Sciences, University of Khartoum, Sudan

* Correspondence:

a.tijjani@cgiar.org, olivier.hanotte@nottingham.ac.uk, hassantahir70@hotmail.com

O. H. and H.M. are joint last co-authors

Abstract

Sudan, the largest country in Africa, acts as a corridor between North and sub-Saharan Africa along the river Niles. It comprises warm arid and semi-arid grazing lands, and it is home to the second-largest African population of indigenous livestock. Indigenous Sudanese cattle are mainly indicine/zebu (humped) type. They thrive in the harshest dryland environments characterised by high temperatures, long seasonal dry periods, nutritional shortages, and vector diseases challenges. We investigated genome diversity in six indigenous African zebu breeds sampled in Sudan (Aryashai, Baggara, Butana, Fulani, Gash, and Kenana). We adopted three genomic scan approaches to identify candidate selective sweeps regions (ZH_p , F_{ST} , XP-EHH). We identified a set of gene-rich selective sweep regions shared across African and Asian zebu or unique to Sudanese zebu. In particular, African and Asian zebu candidate gene-rich regions are detected on chromosomes 2, 5 and 7. They include genes involved in immune response, body size and conformation, and stress response to heat. In addition, a 250 kb selective sweep on chromosome 16 was detected exclusively in five Sudanese zebu populations. This region spans seven genes, including *PLCH2*, *PEX10*, *PRKCZ* and *SKI*, which are involved in alternative adaptive metabolic strategies of insulin signalling, glucose homeostasis, and fat metabolism. Together, these genes may contribute to the zebu cattle resilience to heat, nutritional and water shortages. Our results highlight the putative importance of selection at gene-rich genome regions, which might be under a common regulatory genetic control, as an evolutionary mechanism for rapid adaptation to the complexity of environmental challenges.

Keywords: adaptive divergence, insulin signalling, fat metabolism, desert adaptation, African zebu

Introduction

Sudan is one of the leading African countries in livestock production. The livestock sector plays a critical role in the Sudanese economy and the welfare of the whole population (Wilson, 2018). Of its 101 million livestock heads, cattle comprise approximately 41% (~41 million cattle heads). The remaining are sheep, goats and camels. According to Epstein (1971), cattle were introduced to Sudan from Asia through the Nile Valley and the Horn of Africa. Sudan's main indigenous cattle is the humped *Bos indicus* (zebu) type. The breed names come from human tribes, external morphological traits, specific conformation, size, and branding. The main Sudanese cattle groups include the Northern Sudan short-horned zebu (Kenana, Butana, Baggara), the Nilotic zebu of Southern Sudan (e.g. Toposa-Murle, Mangala) and the Sanga (e.g. Dinka, Nuer). Other cattle populations include White Nile cattle, Fuga or Dar EL Reeh cattle of Northern Kordofan, Nuba Mountains cattle, Gash and Aryashai of East Sudan and the Fulani cattle found in South-West Sudan and across the West Sahelian belt (Rahman 2007, Nahar 2009, HCENR 2014).

Sudan geography offers diversified agro-ecological conditions and climate, differing in rainfall, from as little as 75 mm of annual rainfall in the northern desert to about 1500 mm in the South-West forest. There is also a wide variation in the soil types, temperatures and vegetations. The north part of the country is predominantly desertic, comprising part of the Libyan desert to the West and the Nubian desert to the East, separated by a stretch of the Nile Valley. With virtually no rainfall in this region, the primary water sources that support human settlements and vegetation are a few oases in the Libyan desert. In contrast, the South and West regions consist mainly of sandy plains interrupted by mountains. It extends from the Nuba Mountains to the borders with the Central African Republic and Chad. The Northeastern parts of the country experience little rainfall and are classified as semi-desertic (Impiglia 2017). Therefore, the grazing lands in the arid and semi-arid Sudan regions are characterised by extreme temperatures, high aridity and feed and water scarcity (Musa *et al.* 2006, Ahmed *et al.*, 2016). Undoubtedly, the Sudanese cattle are among the few indigenous African cattle living in extreme dryland climatic conditions. It is expected that such environmental pressures would have left selection footprints in their genomes.

The present study involves whole-genome re-sequencing of six indigenous African zebu breeds sampled in Sudan (Kenana, Butana, Aryashai, Baggara, Gash, and Fulani). Kenana is found in the region between the White and Blue Nile. Butana is located mainly in the Butana plain within a triangle delimited by the River Atbara, the Blue Nile and the River Nile. The natural habitat of Baggara cattle in Sudan is the savannah belt lying between the White Nile and the western fringes of Sudan. It is a breed associated with the Baggara Arabs ethnic group inhabiting the Sahel (mainly between Lake Chad and southern Kordofan). The Baggara nomadic pastoralists in Darfur and Kordofan moved their animals across hundreds of kilometres in search of pastures and water. Aryashai and Gash cattle are mainly found in the El Gash delta, stretching from Aroma in the South, Dordieib in the North and spreading to the Atbara River. Fulani cattle in Sudan are found in the western part of the country, in the Darfur and Blue Nile states (Figure 1). The Falata Umbororo tribes mostly own Fulani cattle in Sudan (Nahar 2009). The breed is also associated with the Fulani people living along the Sahelian belt. In addition, the Kenana and Butana breeds are anecdotally referred to as African dairy zebu because of their superior milk production capability compared to most African zebu breeds. Hence, they are mostly kept under the high input systems of dairy production (Rahman 2007). The production of F1's heifer with exotic taurines for milk production is common across Sudanese cattle, especially among

Genomic signatures of desert adaptation

the dairy Butana and Kenana. Therefore, there is a genuine concern about the ongoing erosion of the genetic diversity of these indigenous breeds under such indiscriminate crossbreeding systems (FAO, 2009; HCENR. 2014, FAO. 2015).

This study assessed the genetic diversity, population structure, and genomic footprints of positive selection in the indigenous Sudanese cattle population. In particular, we aimed to investigate the adaptation of these populations to the extreme dryland environmental condition of the African continent, addressing the question of the genetic origin, within or from outside the African continent, of their ecological adaptations. Accordingly, we applied three genomic scan approaches to detect regions of low within breed diversity (*ZHp*, Rubin *et al.*, 2010), population divergence (F_{ST} , Beaumont 2005) and increased haplotype homozygosity (XP-EHH, Sabeti *et al.*, 2007). We performed these assessments in comparison to other cattle breeds, including two East African zebu breeds (Kenyan Boran, Ethiopian Ogaden), one African sanga breed (Ankole), one Asian zebu from Brazil (Gir), two African taurine breeds (N'Dama, Muturu), three Eurasian breeds (Eastern Finncattle, Western Finncattle and Yakutian cattle) as well two West Europe taurine breeds (Holstein and Angus).

Materials and Methods

Study population and sample re-sequencing

A total of 17 zebu and taurine cattle breeds comprising 185 individuals were included in the study (Table S1). We newly generated the full genome sequences of ninety indigenous Sudanese zebu cattle belonging to two independent datasets of six (Group A, $n = 60$) and three (Group B, $n = 30$) zebu breeds sampled in Sudan. Group A datasets consist of 60 sequences of Kenana (KEN, $n = 10$), Butana (BTN, $n = 10$), Aryashai (ARY, $n = 10$), Baggara (BGR, $n = 10$), Gash (GAS, $n = 10$) and Fulani (FLN, $n = 10$). The Group B datasets consist of 30 sequences of Kenana ($n = 9$), Butana ($n = 11$) and Baggara ($n = 10$). The 30 sequences of the Group B Sudanese datasets were generated more recently as part of the Genomic Reference Resource for African Cattle (GRRFAC) Initiative (<https://grrfac.ilri.org>). Therefore, they were not part of the initial variant calling described below. However, we provide additional Materials and Methods information in Supplementary file one regarding the sampling procedure, read mapping and variant calling involving the Group B dataset.

We also included new sequences of 10 Gir cattle from Brazil (GIR, $n = 10$), provided by Dr Marcos Vinicius Barbosa da Silva of Embrapa Dairy Cattle. The remaining 85 genome sequences are from the public database; Ogaden (OGD $n = 9$, Ethiopia), Kenyan Boran (BOR $n = 10$, Kenya), Savannah Muturu (MUT $n = 10$, Nigeria), N'Dama (NDM $n = 10$, The Gambia), Ankole (ANK, $n = 10$, Uganda), Angus (ANG, $n = 10$), Holstein (HOL, $n = 11$), Eastern Finncattle (EFN $n = 5$) Western Finncattle (WFN $n = 5$) and Yakutian (YKT $n = 5$) (Stothard *et al.* 2011, Bahbahani *et al.* 2017, Kim *et al.* 2017, Weldenegodguad *et al.*, 2019) The NCBI SRA accession numbers are provided in Table S1.

The ethical committee of the Faculty of Veterinary Sciences, University of Nyala, Sudan, approved the sampling protocol for the Group A Sudanese cattle. We obtained 10 ml of whole blood from each animal into EDTA VACUETTE® tubes following the standard procedure under veterinarian supervision. According to the manufacturer protocol, genomic DNA was extracted from the blood samples using the Qiagen DNeasy extraction kit (Qiagen, Valencia, CA, USA). Genomic DNA was evaluated using a NanoDrop spectrophotometer (NanoDrop Technologies, USA) and gel

electrophoresis. Paired-end sequencing was performed on the Illumina® HiSeq platform, with the read length of PE150 bp at each end.

The Gir cattle samples were part of the progeny test program from the National Program for Improvement of Dairy Gir (PNMGL), headed by Embrapa Dairy Cattle (Juiz de Fora, Minas Gerais, Brazil) in cooperation with the Brazilian Association of Dairy Gir Breeders (ABCGIL) and the Brazilian Association of Zebu Breeders (ABCZ). Semen samples were collected for DNA extraction. For this purpose, the semen samples were washed with a lysis buffer and incubated for two hours with an extraction buffer containing dithiothreitol 10% and RNase. Pellets were incubated overnight with a saline-proteinase K buffer, and a phenol-chloroform extraction removed the proteins. The quality and quantity of DNA for all samples were evaluated using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Mate-paired and paired-end libraries (200 bp and 2 x 100 bp, respectively) with different insert sizes were prepared according to the Illumina protocol and subsequently sequenced on the Illumina® HiSeq platform (Illumina Inc., San Diego, CA, USA).

Figure 1. The geography of Sudan shows the sampling areas of the Sudanese zebu cattle populations. Map adapted from Impiglia (2017). Photo credit: Professor Hassan Musa, University of Khartoum, Sudan.

Reads alignment, variants discovery and quality control

We subjected individual sample raw reads to initial quality control using Trimmomatic v0.38 (Bolger *et al.*, 2014). First, we trimmed paired reads of adapter, low-quality bases (qscore < 20) at the beginning and end, then filtered out reads with mean q score less than 20 or length less than 35 bp. The final quality of the resulting clean reads was confirmed using FastQC v0.11.5 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Next, trimmed sequences were aligned to the *Bos taurus* reference genome ARS-UCD1.2 (Rosen *et al.*, 2020), using BWA-MEM v0.7.17 (Li & Durbin, 2010) with default parameters. After mapping, we processed the resultant alignment files and performed variants (SNPs and InDels) discovery following the Genome Analysis Toolkit best practices pipeline (<https://gatk.broadinstitute.org/hc/en-us/sections/360007226651-Best-Practices-Workflows>).

Genomic signatures of desert adaptation

Here, SAMtools *ver* 1.9 was used to convert SAM files to BAM format and for sorting by contigs (Li *et al.*, 2009), duplicates were marked using Picard tools *ver* 2.18.2 (<http://broadinstitute.github.io/picard/>). Variant calling was performed on individual cattle samples using the Haplotype caller of GATK v3.8-1-0-gf15c1c3ef (McKenna *et al.*, 2010) and incorporating known variants from dbSNP *ver*150 (Cunningham *et al.*, 2019). The GATK joint genotyping approach (*GenotypeGVCFs* mode) was adopted to simultaneously identify variants in all cattle samples.

We subjected SNPs and insertions/deletions (InDels) separately to the GATK hard filter (VariantFiltration) steps. The filtering criteria for SNPs include (QD > 2.0, MQ > 40, ReadPosRankSum > 8.0, HaplotypeScore > 13, MappingQualityRankSum > 12.5). The autosomal biallelic SNPs which passed the above filtering criteria and having a Phred-scaled quality score of above 20 (QUAL > 20; approximately 99% likelihood of being correct) were retained for further analyses. The filter criteria for detected InDels include (QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0 || QUAL < 20).

Furthermore, the proportion of missing genotypes in an individual sample and the relatedness among other samples were estimated using the options “--missing-indv” and “--relatedness”, respectively, of Vcftools *v0.1.15* (Danecek *et al.*, 2011). Both individuals of a pair of samples with high relatedness (> 0.8) were excluded from the dataset if they belong to different breeds. In contrast, only one of two individuals with high relatedness was excluded if they belonged to the same breed. Following these criteria, five samples, including two Butana (*n* = 2), two Gash (*n* = 2) and one Baggara (*n* = 1), were removed. No animal was removed due to excessive missing data (> 20%). Hence, 150 cattle samples were retained in downstream analyses.

The total numbers of variants (SNPs and InDels) and the transition to transversion (Ts/Tv) ratio for individual cattle breeds were estimated using BCftools *ver* 1.8 (Li *et al.*, 2009). SnpEff v4.3t (Cingolani *et al.*, 2012) was used to ascertain the genome location and effects of detected variants based on the *Ensembl* cow gene database (ARS-UCD1.2) dbSNP *ver*150. The proportion of detected variants was classified as “known” if the non-reference allele is present in the dbSNP, otherwise as “novel.”

Genomic diversity and inbreeding

From the detected high-quality autosomal SNPs (~ 42.9 M), we estimated the counts of segregating sites for individual cattle using PLINK v2 (Purcell *et al.*, 2007). The average of individuals of the same breed was reported as the per breed estimate. Within-population nucleotide diversity (π) values (Nei and Li 1979) and global averages of pairwise population differentiation (F_{ST}) (Weir and Cockerham 1984) among cattle breeds were estimated based on overlapping 100 kb window and 50 kb step size along bovine autosomes using Vcftools *v0.1.15*. Due to the differences in the number of samples in some of the studied breeds, we randomly selected five representative samples from each breed to estimate these latter parameters.

Population genetic structure

We performed Principal Component Analysis (PCA), admixture, and maximum likelihood tree (Treemix) to investigate the genetic structure among the different cattle breeds using whole-genome autosomal SNPs. Linkage disequilibrium (LD) pruning of SNPs was performed using PLINK default option “50 Kb step 10Kb SNPs, $r^2 > 0.5$ ” (Purcell *et al.*, 2007), resulting in approximately 24 million SNPs. We further used PLINK to generate PCA eigenvectors and eigenvalues in two categories: all 17

195 cattle breeds and Sudanese zebu samples only. PCA plots were generated for the first two eigenvectors
196 using the ggplot2 package in the R *ver* 3.6.3 environment.

197 The block relaxation algorithm implemented in the ADMIXTURE *ver* 1.3.0 software (Alexander et
198 al., 2009) was used to estimate the ancestry proportions of individual cattle samples with kinship (K)
199 set from 2 to 5. Admixture analysis was preceded by removing SNPs with more than 10% missing data
200 (*--geno 0.1*) using VCFtools. Next, we selected the optimal K based on the cross-validation error
201 procedure. Finally, the admixture ancestry ratio was visualised using the pophelper R package (Francis,
202 2017).

203 Phylogenetic relationships based on maximum likelihood tree and possible gene flow events among
204 cattle breeds were investigated using Treemix v1.3 (Pickrell & Pritchard, 2012). The phylogenetic tree
205 was constructed several times, incorporating possible migratory events, from m0 (no migration event)
206 to m10 (ten migration events). Finally, the models were visualised in R using the script provided in
207 Treemix.

208 **Genome diversity analysis and functional annotation**

209 We applied three genomic scan approaches to detect regions of low within breed diversity (*ZHp*),
210 population differentiation (F_{ST}) and increased haplotype homozygosity (XP-EHH) as a proxy of signals
211 for candidate selective sweep in the indigenous African zebu from Sudan.

212 We applied the within-population *Hp* test to detect genomic signatures of low diversity in each of the
213 eight African zebu breeds and the Gir breed. The *Hp* analysis involves counting reads with the most
214 and most minor abundantly observed alleles at every SNP position within a specified window size and
215 sliding step size. The distribution of *Hp* values was normalised by transforming *Hp* to Z-scores (*ZHp*)
216 using ($ZHp = (Hp - \mu Hp) / \sigma Hp$) (Rubin *et al.*, 2010, Rubin *et al.*, 2012, Axelsson *et al.*, 2013). Using
217 VCFtools (Danecek *et al.*, 2011), we estimated regions of population divergence (F_{ST}) based on Weir
218 and Cockerham (1984) between the Sudanese zebu populations and the combined 11 zebu and taurine
219 breeds (Gir, Ogaden, Kenya Boran, Ankole, Muturu, N'Dama, Angus, Holstein, Western Finncattle,
220 Eastern Finncattle and Yakutian). We performed two genome-wide F_{ST} analyses using independent
221 datasets of Sudanese zebu populations; Groups A (Arayshai, Baggara, Butana, Gash, Fulani and
222 Kenana) and B (Baggara, Butana and Kenana). The weighted F_{ST} were also Z-transformed in R. In
223 each *ZHp* and F_{ST} test, we estimated genome-wide test statistics using 100 kb windows and sliding 50
224 kb step across the bovine 29 autosomes and excluded windows containing less than 10 SNPs. Also, the
225 100-kb windows in the extreme top 0.5% test values were arbitrarily defined as candidate selection
226 outlier windows. We find an overlap of detected selective sweep regions between the two F_{ST} analyses
227 using BEDTools intersect (*version* 0.2.29) (Quinlan and Hall, 2010).

228 *Bos taurus* (ARS-UCD1.2) genes overlapping the candidate selected windows were retrieved based on
229 the *Ensembl* cow genes database 104 using the *Ensembl* BioMart online tool
230 (<http://www.ensembl.org/biomart>) (Smedley *et al.*, 2009). Then, we generated a Venn diagram based
231 on the list of detected selective sweep regions based on *ZHp* and F_{ST} using the online tool at
232 (<https://molbiotools.com/listcompare.html>).

233 Finally, following the *ZHp* and F_{ST} analysis results, we further investigate the candidate selective
234 sweep on chromosome 16 uniquely detected in five of the six Sudanese zebu breeds by performing the
235 LD-based XP-EHH test using the Rehh *ver* 2.0 R package (Gautier *et al.*, 2017). Here, we contrasted
236 the combined Group A Sudanese zebu breeds (except Fulani) against the Fulani, the other zebu breeds
237 (Ethiopia Ogaden, Kenya Boran and Gir), and the taurine breeds (African and European taurine). As

Genomic signatures of desert adaptation

this analysis required phased haplotype, we performed phasing of the entire chromosome 16 SNP data for all 155 cattle samples (excluding Group B Sudanese sequences) using the default parameter of Beagle ver 5.0 (Browning *et al.*, 2007), except for N_e , which was set to 1000 to improve the accuracy of phase information (Dutta *et al.*, 2020)

Results

Sequence reads and variants statistics

After removing adapter sequences and low-quality reads, we mapped individual sequences to the *Bos taurus* genome of reference, ARS-UCD1.2 (Rosen *et al.* 2020). We achieved an average alignment rate of up to 99% in all samples. Approximately 43 million SNPs were detected in all the 150 cattle samples. However, the number of segregating sites (homozygous and heterozygous alternate SNPs) in each breed ranges from more than 23 million in the zebu (humped) breeds to between 9 and 14 million in the taurine (humpless) breeds. Within individual humped cattle, the average number of SNPs ranged from around nine million in Kenya Boran to 12 million in Kenana and Butana. On the other hand, a range of 4.5 million to 6.3 million SNPs was detected in the humpless African and European cattle. Therefore, the number of SNPs identified in individual humped cattle is approximately 2 – 2.5 times higher than the number of SNPs in individual taurine cattle (Table S2).

The estimated heterozygous to homozygous SNPs in most 17 cattle breeds is more than 1. The exceptions being the Gir (Het/Hom = 0.74), the Muturu (Het/Hom = 0.63) and the Eastern Finncattle (Het/Hom = 0.86). The highest ratio among the Sudanese zebu breeds is in Butana (Het/Hom = 1.53), while the lowest is observed in Gash (Het/Hom = 1.01). The average transition versus transversion (Ts/Tv) ratio for the 17 cattle breeds is above 2.3 (Table S2).

The numbers of small insertions and deletions (InDels) in individual humped cattle range from about 1.2 million in most African and Asian breeds to approximately 1.5 million in Kenana and Butana. In comparison, the number of InDels in the taurine individuals varies between 600,000 to 800,000 (Table S2). In contrast to the database of known cattle variants (dbSNP ver150, last accessed December 2020), an average of 1.7 M (~ 6.5%) novel SNPs and 1.7 M (~5 0.5%) novel InDels were detected for the Sudanese population (Table S2). These represent a valuable addition to the database of known cattle variants.

Genetic diversity and population differentiation

The average nucleotides diversities among 17 cattle breeds studied are the highest in the eight African zebu breeds (between 2.63×10^{-3} and 3.2×10^{-3}), followed by the Gir (2.49×10^{-3}), the African sanga, Ankole (2.20×10^{-3}), and the taurine breeds (between 1.50×10^{-3} to 1.24×10^{-3}). The lowest value is observed in Muturu (9.94×10^{-4}) (Figure 2). On the other hand, nucleotide diversities are the highest within the African zebu breeds in the two African zebu dairy breeds, Butana and Kenana (~ 3.2×10^{-3}) (Figure 2A).

Population differentiation (F_{ST}) is low among the African zebu breeds (Table S3). The average highest population divergence is between the Sudanese Gash and the taurine breeds ($F_{ST} = 0.324$). It is the lowest between the Sudanese Baggara and the taurine breeds ($F_{ST} = 0.267$). Interestingly, the average population divergence between the African zebu and the West African taurine is higher for the West

278 African shorthorn, Muturu ($F_{ST} = 0.342$), compared to the N'Dama, a West African longhorn ($F_{ST} =$
279 0.245) (Table S3). Population divergence between each of the African zebu and the Gir cattle is around
280 0.1.

281

282 **Figure 2.** Genetic diversity and population structure of the 17 studied breeds. (A) Boxplot of average
283 nucleotide diversities (π). The nucleotide diversity level within each cattle breed was calculated based
284 on an overlapping 100 kb window with a 50 kb step size. (B) Principal component (PC 1 versus PC
285 2) analysis for 17 cattle breeds. (C) Principal component (PC 1 versus PC 2) analysis for the six
286 Sudanese zebu cattle populations. (D) Admixture plot showing ancestry proportions for the 17 cattle
287 breeds. The population structure was assessed using ADMIXTURE ver.1.3.0. The individual
288 population are represented by a vertical bar and partitioned into coloured segments. Each segment's
289 length represents the proportion of the inferred number of ancestries ($K = 2$ to $K = 5$). GIR – Gir, GAS
290 – Gash, ARY – Aryashai, BUT – Butana, KEN – Kenana, BGR – Baggara, FLN – Fulani, OGD –
291 Ogaden, BOR – Kenya Boran, ANK – Ankole, NDA – N'Dama, MUT – Muturu, YKT – Yakutian,
292 WES – Western Finncattle, EAS – Eastern Finncattle, AAN – Angus and HOL – Holstein. The cattle
293 breeds were also grouped into six entities, namely, Asian zebu (ASZ), Sudanese zebu (SUD), East
294 African zebu (AFZ), African sanga (AFS), West African taurine (WAT) and European taurine (EUT).

295 Population structure

296 The principal component analysis of the 17 cattle breeds was performed using approximately 24
297 million autosomal bi-allelic SNPs after removing SNPs with a minor allele frequency of less than 0.05.
298 The first and second principal components account for 54.9% and 12.2% variations. PC 1 clearly
299 distinguished the taurine from the zebu breeds, while PC 2 separates African breeds from non-African
300 breeds (Figure 2B). A second PCA, including only the Sudanese zebu, reveals pairs of clusters (Kenana
301 - Butana, Aryashai - Gash, and Baggara – Fulani), with no clear differentiation between the Baggara
302 and the Fulani. PC 2 (12.1%) shows the separation of Kenana and Butana (African zebu dairy breeds)
303 from the other Sudanese zebu (Figure 2C).

304

305 After LD pruning ($r^2 > 0.5$), a dataset of approximately 4.5 million SNPs was used to explore the
306 ancestry proportions (K) of the different cattle breeds for K ranging from 2 - 5 (Figure 2D). The

Genomic signatures of desert adaptation

admixture cross-validation (CV) option (Alexander *et al.* 2009) supports $K = 4$ with the lowest CV error as the likely optimum number of ancestral backgrounds for our populations. The four possible ancestral backgrounds inferred here are *Bos indicus*, European *Bos taurus*, and two distinct African *Bos taurus* ancestries (N'Dama and Muturu). At this level, the African zebu breeds show admixed zebu and taurine genetic backgrounds, as in previous reports (Kim *et al.*, 2020), including two distinct African *Bos taurus* ancestries, the longhorn N'Dama and the shorthorn Muturu (Tijjani *et al.*, 2019). On the other hand, a low level of European taurine ancestral background is observed in the Sudanese Baggara, Fulani and the two East African zebu breeds (Figure 2D). Furthermore, at five ancestral backgrounds ($K = 5$) reveal two separate African zebu ancestries, whereby the dryland Sudanese zebu are distinct from the East African zebu breeds.

Both PC analyses (Figure 2C) and admixture analysis (Figure 2D, $K = 2$ to 5) indicate substantial genetic similarity between the Sudanese Baggara and the trans-Saharan Fulani, with no separation between the two populations in the PC analysis and very similar ancestral background (admixture analysis).

Figure 3. – Maximum likelihood evolutionary tree and possible gene flow among the 17 cattle breeds: (A) without migration events, (B) assuming ten migration events. E-Finn Eastern Finncattle) W-Finn r(Western Finncattle), K-Boran (Kenyan Boran).

Evolutionary relationships among cattle breeds

As for the admixture analyses, genome-wide unlinked autosomal SNPs were used for the Treemix analyses. The phylogenetic relationships and migration events reveal several possible genes flows among the cattle breeds (Figure 3). By sequentially adding up to 10 migration events and agreeing with the admixture analysis, we observe possible gene flow from the African taurine into most African zebu breeds, including four Sudanese populations (Kenana, Butana, Baggara and Fulani). On the other hand, we did not observe gene flow between the Sudanese zebu and any European breeds. Our Treemix results also reveal possible introgression between the two West African taurine breeds with the direction of gene flow from the shorthorn Muturu to the longhorn N'Dama (Figure 3).

Genomic signatures of positive selection in zebu populations

We performed a genome-wide autosomal *Hp* selection scan in the nine zebu breeds of six Sudanese zebu populations, Kenyan Boran, Ogaden and Gir. The distributions of Z-transformed *Hp* scores across the bovine autosome are shown in Figure 4 (Sudanese populations) and Figure S1 (other zebu breeds). We considered 249 outlier windows (100-kb windows in the lowest 0.5% of *ZHp* scores) in each breed as regions with significantly reduced diversity and possible signatures of positive selection (Tables S4 – S12). The highest *ZHp* threshold is observed in Baggara (*ZHp* = -3.251), while the lowest is observed in Gir (*ZHp* = -3.958) (Figures 4 and S1). We identify 273 outlier windows common to at least two breeds among the Sudanese zebu population. We merged the windows in the proximity of up to 5 kb, resulting in 114 selective sweeps regions varying in size from 0.1 to 1.5 Mb. Fifty-five selective sweep regions overlap with outlier windows detected in at least one non-Sudanese zebu population. The remaining 59 selective sweep regions were detected exclusively in the Sudanese zebu populations. Thus, these regions are shared and unique Sudanese zebu selective sweep regions (Table S13). Based on *Ensembl* cow genes 104 (ARS-UCD1.2), 47 out of the 55 shared Sudanese – non-Sudanese zebu selective sweep regions overlap with protein-coding genes ($n = 227$) (Tables S14). In contrast, of the 59 unique Sudanese selective sweeps, 40 overlaps with 135 protein-coding genes (Tables S15). In both cases, more than 60 per cent of the annotated regions contain more than one protein-coding gene (gene-rich). However, the selective sweep regions devoid of any annotated cow genes (desert gene regions) spans a total 4.15 Mb ARS-UCD1.2 genome region (Table S16)

Figure 4. Genome-wide distribution of *ZHp* scores across bovine autosomes in the six Sudanese zebu populations. The blue line indicates the *ZHp* threshold value (lowest 0.5%) for selecting outlier windows (candidate regions under positive selection).

Genomic signatures of desert adaptation

Shared African and Asian zebu selective sweep regions

Among the 47 shared selective sweep regions overlapping annotated genes, two regions, on chromosome 5 (5:47.40 – 48.0 Mb) and 7 (7:49.85 – 51.15 Mb), are detected in eight of the nine zebu breeds. We detect the 600 kb region on chromosome 5 in all the six Sudanese zebu populations, the Kenyan Boran and Gir breeds but not in the Ethiopian Ogaden. The region overlaps eight genes (*ENSBTAG00000052954*, *ENSBTAG00000053419*, *GRIPI1*, *HELB*, *HMGA2*, *IRAK3*, *LLPH*, *TMBIM4*) with genes function related to growth, conformation, reproduction and the immune systems (Porto-Neto *et al.*, 2014, Naval-Sánchez *et al.*, 2020). The approximately ~1.5 Mb gene-rich sweep on chromosome 7 overlaps with up to 28 protein-coding genes, including two heat shock proteins, *DNAJC18* and *HSPA9*. Again, this large sweep is detected in eight zebu breeds, except the Sudanese Butana. Other shared selective sweep regions between African and Asian breeds (present in at least six breeds) are found on chromosomes 2 (2:70.25 - 70.50 Mb), 8 (8:59.35 - 59.70 Mb), 10 (10:58.85 - 59.15 Mb) and 17 (17:13.15 - 13.35 Mb) (Table S14).

We found several candidates selected regions in African zebu breeds which are not present in the Asian Gir. These include regions on chromosomes 7 (7:52.10 – 52.40 Mb), 11 (11:13.10 – 13.30 Mb), and 20 (20:48.80 – 48.95 Mb) detected in seven of the eight African breeds. The exceptions are Butana (chromosome 7), Gash (chromosome 11) and Kenana (chromosome 20) (Table S14). In particular, the region on chromosome 7 overlaps with up to eight protein-coding genes (*ENSBTAG00000030474*, *ENSBTAG00000049316*, *ENSBTAG00000049493*, *ENSBTAG00000050840*, *ENSBTAG00000053160*, *PCDHAC2*, *PCDHB1* and *SLC25A2*). Also, we found another gene-rich region on chromosome 19 (19:26.35 – 26.50 Mb), containing up to fourteen protein-coding genes detected in six African zebu breeds, including four Sudanese populations (Table S14).

Besides, we identified regions in the Sudanese zebu population shared only with the Gir. These include a 650 kb region on chromosome 5 (5:48.4 - 49.05 Mb) found in five Sudanese zebu populations, the exception being the Baggara. This region overlaps five genes (*ENSBTAG00000000237*, *LEMD3*, *MSRB3*, *TBC1D30* and *WIFI1*). Also, a 150 kb candidate selected region on chromosome 2 is detected in four Sudanese populations and the Gir. It overlaps with the macrophage receptor with the collagenous structure gene (*MARCO*), involved in the innate immune response (Kissick *et al.*, 2014). Additionally, a 150 kb region on chromosome 5, overlapping with the *HOXC* gene cluster, was also detected exclusively in four Sudanese zebu populations and the Gir (Table S14). This region has also been reported previously in the Gir cattle (Liao *et al.*, 2013).

Sudanese zebu specific candidate selective sweeps

We classified as Sudanese zebu-specific sweeps 59 genome regions of low diversity (*ZHp*) detected uniquely in the Sudanese zebu populations. The commonest sweep regions are a 450 kb region on chromosome 7 (7:54.0 – 54.45 Mb) and a 250 kb region on chromosome 16 (16: 50.40 – 50.65 Mb). They are present in five Sudanese breeds, the exception being Baggara and Fulani, respectively (Table S15). The region on chromosome 7 contains *ARHGAP26* and *NR3C1*, while chromosome 16 region contains seven genes (*FAAP20*, *MORN1*, *PEX10*, *PLCH2*, *PRKCZ*, *RER1*, and *SKI*). In addition, we identified in four Sudanese breeds a 200 kb region on chromosome 7 (7:19.85 – 20.05 Mb). This region is downstream of another 150 kb region (7:19.60 – 19.75 Mb) detected in three Sudanese populations. The two regions are close to each other, being only separated by 100 kb. Together, they overlap with 16 genes (Table S15). More so, a gene-rich region of 200 kb on chromosome 5 (5: 55.85 - 56.05 Mb) was detected in three populations. It includes seventeen genes (*ARHGAP9*, *ARHGEF25*, *B4GALNT1*,

DCTN2, *DDIT3*, *DTX3*, *ENSBTAG00000049386*, *ENSBTAG00000051574*, *ENSBTAG00000051593*, *GLI1*, *INHBC*, *INHBE*, *KIF5A*, *MARS1*, *MBD6*, *PIP4K2C*, and *SLC26A10*). The remaining regions overlapping at least one protein-coding gene are detected only in two Sudanese zebu populations (Table S15).

Genetic differentiation at shared and unique Sudanese zebu candidate selected genome regions

We searched for selective sweeps possibly linked to environmental adaptation in the arid and semi-arid African regions. Hence, we looked for evidence of divergence at genome regions between the Sudanese zebu population and the other studied cattle populations following the Weir and Cockerham (1984) genetic differentiation index (F_{ST}). Firstly, we calculated F_{ST} values between the combined six Sudanese zebu populations and the combined 11 non-Sudanese zebu and taurine breeds (see Materials and Methods). The genome-wide distribution of Z-transformed F_{ST} values across the 29 bovine autosomes is presented in Figure 5A. We identified 106 outliers 100-kb windows in the top 0.5 % positive ZF_{ST} values ($0.3 \leq F_{ST} \leq 0.5$ and $4.1 \leq ZF_{ST} \leq 7.4$). Neighbouring windows were merged, resulting in 33 selective sweeps spanning 149 protein-coding genes (Table S17).

Next, we repeated the genome-wide F_{ST} analysis with an independent dataset of Sudanese zebu cattle genome sequences to further confirm the selected regions of population divergence between the indigenous Sudanese zebu population and other non-Sudanese cattle populations. We refer to the initial Sudanese genome dataset as Group A and the new genome dataset as Group B. The new dataset comprises 30 newly generated genome sequences of Baggara ($n = 10$), Butana ($n = 11$) and Kenana ($n = 9$) from the Genomic Reference Resource for Africa cattle (GRRFAC) initiative. Following a similar F_{ST} approach with the Group A dataset, we identified 105 outliers 100-kb windows within the top 0.5 % positive ZF_{ST} values ($\sim \geq 0.3 F_{ST} \leq 0.5$ and $\geq 3.9 ZF_{ST} \leq 7.25$). These windows were merged into 36 selective sweep regions overlapping 154 protein-coding genes (Figure 5B, Table S18).

By comparing the two F_{ST} results, we found an overlap of 101 protein-coding, positively selected genes clustered into eighteen mostly gene-rich selective sweep regions (Table S19). Among them are 11 selective sweeps regions classified as shared ($n = 8$) and unique ($n = 3$) Sudanese zebu selective sweep spanning 41 and 23 positive selected genes, respectively (Figure 5C, Table 1). Of the eight highly differentiated, shared selective sweep regions, four are found on chromosome 5 ($n = 4$). In addition, the strongest F_{ST} signals in the two analyses (mean $Z_{F_{ST}} = 7.18$) are found on one of the regions on chromosome 5 (5:47.25 – 47.95) (Figure 5A-B). The remaining highly differentiated and shared selective sweeps are found on chromosomes 2 ($n = 1$), 7 ($n = 2$) and 8 ($n = 1$) (Tables 1). The two regions on chromosome 7 add up to 0.8 Mb and are part of the ZHp detected 1.5 Mb region (Table S14)

The 23 Sudanese zebu-specific highly differentiated, positive selected genes are distributed across three selective sweep regions on chromosomes 5, 7 and 16. Each of these regions spans 11, 5 and 7 protein-coding genes, respectively (Table 1). Intriguingly, at least one gene within these regions is directly or indirectly involved in insulin signalling and glucose metabolism, which are likely relevant adaptive metabolic processes in hot and arid environments (Rocha *et al.*, 2021). Examples of these genes are *KIF5A* (Cui *et al.*, 2011), *PIP4K2C* (Shim *et al.*, 2016), *MAP2K2* (Laskowski *et al.*, 2016), *FAAP20* (Lagundžin *et al.*, 2019), *PRKCZ* (Zou *et al.*, 2013, Ahmed *et al.*, 2020) and *SKI* (Diaz *et al.*, 2012). However, while the chromosomes 5 and 7 unique selective sweeps are detected in fewer Sudanese zebu breeds (≤ 4) than the chromosome 16 region detected exclusively in five Sudanese breeds (except the Fulani). Also, the chromosome 16 region shows the strongest average F_{ST} signal (Tables S15 and 1). Likewise, this region is highly differentiated between the five Sudanese zebu and

Genomic signatures of desert adaptation

the taurine populations (combined European and African taurine breeds) ($F_{ST} > \sim 0.75$), non-Sudanese zebu populations (East African and Asian zebu) and Sudanese Fulani ($\sim 0.1 F_{ST} \leq 0.25$) (Figure 6B). Hence, we opted to investigate this particular region further.

Figure 5 – Genome-wide distribution of ZF_{ST} scores along *Bos taurus* (ARS-UCD1.2) autosomes following the population differentiation F_{ST} analyses between each of two different datasets of Sudanese zebu populations (A). Group A, $n = 60$ and (B). Group B, $n = 30$, combined eleven other zebu and taurine cattle breeds. The Group A populations consist of six Sudanese breeds, namely; Aryashia, Baggara, Butana, Fulani, Gash and Kanana. Group B samples consist of Baggara, Butana and Kenana breeds. The blue lines indicate the top 0.5% ZF_{ST} values threshold to consider outlier regions. (C). Venn diagram showing the numbers of overlapping detected protein-coding genes between ZHp and F_{ST} tests. The F_{ST} detected genes are the 101 common genes following the two F_{ST} analyses involving group A and group B Sudanese zebu population

A unique Sudanese zebu-specific candidate selective sweep at chromosome 16 locus

To further confirm the evidence of recent selection at the chromosome 16 locus and possibly avoid the problem of confounding demographic factors (Oleksyk *et al.*, 2010), we performed additional genomic scan analysis on the entire bovine chromosome 16 using the XP-EHH test (Sabeti *et al.*, 2007). We used the Sudanese Fulani, the other zebu breeds and the taurine population as three separate control groups (Figure S2A-C). The haplotype homozygosity in the Sudanese zebu population on the chromosome 16 locus is shown in Figure 6C. In particular, the comparison with Fulani reveals selection signals at four significant SNPs ($-\log_{10} XP-EHH, > 4$), which are located within three genes (*PRKCZ* (16:50607838 and rs720151551), *SKI* (rs716750470) and *PLCH2* (rs525024181) (Figure 6D). In contrast, XP-EHH comparison with other zebu populations (East African and Asian breeds) reveals one strong signal containing several significant SNPs (Figure 6E). The latter's top four significant SNPs (rs797777164, rs519790637, rs525024181 and rs523418636) are located within the *PLCH2* gene.

Consequently, *PRKCZ*, *SKI* and *PLCH2* genes might be zebu-specific candidate genes linked to local

Genomic signatures of desert adaptation

adaptation to the dryland habitat in Sudan's desert and semi-desertic regions. Indeed, they are involved in alternative metabolic processes of glucose homeostasis, insulin signalling and fat metabolism, which are likely relevant adaptive strategies in hot and arid environments (Diaz *et al.*, 2012, Zou *et al.*, 2013, Lagundzin *et al.*, 2019, Ahmed *et al.*, 2020, Rocha *et al.*, 2021). Furthermore, we considered the one common significant SNP (rs525024181, G>A) within the *PLCH2* genes (Figure 6D-E) as a putative candidate variant or closely linked polymorphism targeted by selection. The haplotype decay around rs525024181 reveals that the haplotype containing the derived allele has more extensive homozygosity in the Sudanese zebu than in the other zebu breeds (Figure 6F - H), suggesting a strong selection at the locus in the Sudanese population. We did not observe much difference in the pattern of EHH around the remaining seven significant SNPs among the zebu populations (Figure S3A - C).

Figure 6 - Strong evidence of positive selection in Sudanese zebu population at 250-kb chromosome 16 locus. (A) Genomic footprints of low diversity (ZHp) in the candidate region in eight African zebu breeds and one Asian zebu breed. The red and black dashed line indicates the maximum and minimum genome-wide ZHp selection outlier threshold values among the cattle breeds. (B) Population divergence (F_{ST}) between the Sudanese zebu populations (excluding the Fulani) and other zebu and taurine cattle populations. (C) Compared to other cattle populations, there was increased haplotype homozygosity (XP-EHH) at a 250 kb selective sweep locus in Sudanese zebu. (D, E) Localisation of SNPs at the selective sweep region following XP-EHH between the combined five Sudanese breeds and Fulani (D), as well as other zebu breeds (East African and Asian breeds) (E). (F, G, H). The decay of extended haplotype homozygosity around rs525024181 in Sudanese Zebu (F), Fulani (G) and other zebu breeds (H).

Table 1. Examples of shared with non-Sudanese and Sudanese specific gene-rich and high population divergent selection signatures of putatively linked adaptive phenotypes for the dryland environment.

Genomic signatures of desert adaptation

Chr	Start ^{\$}	End ^{\$}	ZHp	F _{ST} [*]	ZF _{ST}	Overlapping <i>Bos taurus</i> genes [^]
2	70.60	70.80	Shared	0.42	5.02	<i>EN1, MARCO</i>
5	25.85	26.10	Shared	0.47	5.87	<i>HOXC10,HOXC11,HOXC12,HOXC13,HOXC4,HOXC5,HOXC6,HOXC8,HOXC9,SMUG1</i>
5	47.25	47.95	Shared	0.55	7.18	<i>ENSBTAG00000052954,ENSBTAG00000053419,GRIP1,HELB,HMGA2,IRAK3,LLPH,TMBIM4</i>
5	48.35	48.85	Shared	0.50	6.38	<i>LEMD3,MSRB3,WIF1</i>
5	55.85	55.95	Unique	0.30	4.2	<i>ARHGEF25, B4GALNT1, DCTN2, DTX3, ENSBTAG00000049386, ENSBTAG00000051574, ENSBTAG00000051593, KIF5A, MBD6, PIP4K2C, SLC26A10</i>
5	56.90	57.10	Shared	0.40	4.76	<i>ANKRD52,APOF,APON,CNPY2,COQ10A,CS,ENSBTAG00000052361,GLS2,IL23A,MIP,NABP2,PAN2,SLC39A5,SPRYD4,STAT2,TIMELESS</i>
7	19.75	20.00	Unique	0.40	4.5	<i>CREB3L3,MAP2K2,PIAS4,SIRT6,STAP2,ZBTB7A</i>
7	50.00	50.20	Shared	0.39	4.49	<i>CTNNA1,ENSBTAG00000004415,LRRTM2</i>
7	50.60	51.10	Shared	0.41	4.86	<i>CXXC5,DNAJC18,ECSCR,MZB1,NRG2,PROB1,PSD2,SLC23A1,SMIM33,SPATA24,STING1,UBE2D2</i>
8	57.80	58.05	Shared	0.46	5.78	<i>ENSBTAG00000005667</i>
16	50.40	50.65	Unique	0.45	5.58	<i>FAAP20, MORN1, PEX10, PLCH2, PRKCZ, RERI, SKI</i>

^{\$}Unit is in megabases (Mb)

^{*}population differentiation (F_{ST}) between combined Sudanese zebu breeds and the other eleven zebu and taurine cattle breeds studied.

[^]Positive selected genes based on the overlap between ZHp and F_{ST} genomic scan tests in Sudanese zebu breeds.

Discussions

In this study, we analyse for the first time the whole-genome sequences of six indigenous African zebu breeds from the most extreme African dryland conditions of Sudan. We aimed to identify candidate selective sweep regions linked to dryland adaptive traits and to assess their uniqueness among other African and non-African cattle. We adopted three genomic scan approaches to detect genomic regions of reduced diversity (ZHp), population divergence (F_{ST}), and increased haplotype homozygosity (XP-EHH). Most of the within-breed ZHp genomic signatures detected are unique to individual zebu breeds. Interestingly, we find several gene-rich regions among the shared candidate selective sweeps across Sudanese or between Sudanese and other zebu breeds. In addition, some of these regions also show high population divergence signals between the Sudanese zebu and other zebu and taurine breeds. These results are compatible with the identified selective sweeps being *Bos indicus* specific, while others being rather African dryland specifics as detected exclusively in the Sudanese cattle population of the arid environments. If under common regulatory control, these gene-rich regions will be consistent with the view that change in gene expression may control complex traits (Naval-Sanchez *et al.*, 2018) while providing a possible mechanism for rapid adaptation to new environmental challenges (Gheyas *et al.* 2021)

Interestingly, the functions of the genes within candidate gene-rich regions identified in the present study are associated with phenotypes such as body size and conformation, stress response to heat, immune response, insulin signalling, glucose metabolism, and fat metabolism. Interestingly, positively selected genes involved in similar processes have been reported in other desert-dwelling mammals like sheep, goats and camels. However, many reported genes are not the same (Wu *et al.*, 2014, Kim *et al.*, 2016, Mwacharo *et al.*, 2017). Nevertheless, the results here support the evidence of common adaptive traits or mechanisms among the ungulates adapted to hot and arid climates (Mirkena *et al.*, 2010, Gebreyohanes *et al.*, 2017, Rocha *et al.*, 2021).

Among the highly frequent *Bos indicus*-specific, selective sweeps are candidate genes involved in body size, conformation, stress response to heat and immune response. Notably, the *HMGA2* gene within the expanded selective sweep on chromosome 5 has been linked to different phenotypes, including growth and conformation in other non-African cattle breeds (Porto-Neto *et al.*, 2014, Yurchenko *et al.*, 2018, Aguiar *et al.*, 2018, Naval-Sánchez *et al.*, 2020, Dixit *et al.*, 2021), and other species (Kim *et al.*, 2004, Weedon *et al.*, 2007, Boyko *et al.*, 2010, Fariello *et al.*, 2014, Sevane *et al.*, 2017 and Dutta *et al.*, 2020). Interestingly, for the first time, we identified strong selection signatures in the region of this pleiotropic gene in the indigenous African cattle population. On the contrary, several genes overlapping the extended gene-rich selective sweep on chromosome 7 have previously been detected in African and non-African zebu breeds (Kim *et al.*, 2017, Bahbahani *et al.*, 2017, Kim *et al.*, 2020, Naval-Sánchez *et al.*, 2020). In particular, Kim *et al.* (2020) reported an excess of indicine ancestry in this region. Furthermore, among the twenty-eight genes in this region is the DnaJ heat shock protein family (HSP40) member C18 (*DNAJC18*), also reported in tropically adapted sheep breeds from Ethiopia (Ahbara *et al.*, 2019) and Bactrian camel (Wu *et al.*, 2014). Undoubtedly, the genes in this region are involved in stress response to heat. However, the region requires further investigation and fine mapping to identify the probable candidate gene(s) or variant(s) targeted by selection.

Exposure to infectious and parasitic diseases challenges is another stressor faced by Sudanese cattle living in hot arid environments, necessitating an adequate immune response for survival in the region. We found the interleukin 1 receptor-associated kinase 3 (*IRAK3*) gene in the proximity of the *HMGA2* gene. *IRAK3* is a crucial modulator of inflammatory responses and negatively regulates Toll-like receptor signalling pathways involved in innate host defence (Nguyen *et al.*, 2020). It has been reported under positive selection in non-African humped cattle adapted to tropical climatic conditions (Porto-Neto *et al.*, 2014; Naval-Sánchez *et al.*, 2020). Another detected innate immune-related gene is the macrophage receptor with collagenous structure (*MARCO*), shared by most Sudanese breeds and the Gir cattle but not the African non-Sudanese zebu breeds. The gene is likely involved in immune responses against bacterial infection, particularly the causative agent of tuberculosis, *Mycobacterium bovis*. Specifically, *MARCO* plays a role in the phagocytosis of bacteria, including *Mycobacterium* species and *Salmonella* species, as demonstrated in previous studies involving microbial infections in mice and zebrafish (Dorrington *et al.*, 2013; Benard *et al.*, 2014). Polymorphisms in the gene have also been associated with pulmonary tuberculosis (TB) in humans caused by *Mycobacterium tuberculosis* (Ma *et al.*, 2011; Bowdish *et al.*, 2013; Thuong *et al.*, 2016). Although *M. tuberculosis* is the causative agent of TB in humans, the incidence of this anthropogenic disease is frequently identified in cattle, particularly in Africa, including Sudan, where the disease's prevalence is also high in humans. It is believed that the transmission between man and cattle results from the close association of cattle farmers to their animals (Ocepek *et al.*, 2005; Thoen *et al.*, 2008). Indeed, Sudan happens to be one of Africa's regions where a very high prevalence of TB in cattle herds has been reported (Sulieman and Hamid 2002; Ibrahim *et al.*, 2016). Also, the only two anecdotally known African dairy zebu breeds, Butana and Kenana, are indigenous to Sudan. This gene's role in TB susceptibility in humans and its

Genomic signatures of desert adaptation

detection in this study in a candidate region under positive selection in African cattle further supports cattle as asymptomatic carriers of the human's *Mycobacterium tuberculosis*.

In addition to extreme temperature and disease challenges, other harsh conditions of the African dryland areas include food and water shortages (Jonsson 2006, Uddin and Kebreab 2020). Indeed, we identified a unique Sudanese zebu selective sweep on chromosome 16 detected exclusively in five Sudanese breeds (except the Fulani). Furthermore, this region shows high population differentiation between the Sudanese zebu and other cattle populations, including the Sudanese Fulani. The absence of this selection signal in the Fulani will need to be further investigated, considering the genetic closeness of the Fulani and Baggara and the origin of the Fulani breed, predominantly found in the West African Sahelian region (Rahman 2007, Nahar 2009). The candidate genes overlapping this unique sweep have roles in glucose homeostasis, feed efficiency, insulin signalling and fat metabolism. Glucose homeostasis and insulin signalling are essentially biological processes contributing to animals' survival due to nutritional shortages in the desert. Also, there are pieces of evidence of the periods of dietary restriction in cattle and other species coinciding with increased insulin sensitivity and glucose uptake (Cartee *et al.*, 1994, Röpke *et al.*, 1994, Sternbauer and Luthman 2002, Barretero-Hernandez *et al.*, 2010). In addition, *FAAP20*, *PRKCZ*, and *SKI* genes have been implicated in insulin metabolism (Leong *et al.*, 2010, Diaz *et al.*, 2012, Fitzsimons *et al.*, 2014, Kriebel *et al.*, 2016, Lagundžin *et al.*, 2019). In particular, *PRKCZ* functions by regulating the translocation of the glucose transporter 4 (*GLUT4*) to the cell surface for glucose uptake and has been implicated in insulin response to glucose uptake in fasted cattle. Thus, a greater expression of both *PRKCZ* and *GLUT4* has been observed in fasted animals, suggesting a greater sensitivity to glucose uptake by myocytes (Keogh *et al.*, 2015).

Interestingly, consistent with previous reports on the profound increased expression levels of *GLUT1* (glucose transporter 1) and genes involved in glycolysis in the renal medulla of water-deprived Bactrian camels (Wu *et al.*, 2014). Thus, our results suggest shared physiological responses facilitated by glucose transporter proteins in camelids and cattle during water and food scarcity. However, the processes may have evolved more rapidly in camelids than in cattle as a result of *GLUT1* being ubiquitously expressed and possessing high capacity than the *GLUT4*, which is expressed mainly in insulin-sensitive tissues like the heart, skeletal muscle and adipose tissue (Scheepers *et al.*, 2004, Wu *et al.*, 2014). In addition, increased expression of the insulin-responsive glucose transport proteins has been correlated with improved insulin action in skeletal muscle, especially during exercise. Indeed, overexpression of the *SKI* gene in the skeletal muscle has been demonstrated to modulate the genetic controls of insulin signalling and, thus, glucose homeostasis (Stallknecht *et al.*, 2000, Diaz *et al.*, 2012).

The above function may correlate with whole-body fat reduction in exercised humans due to increased insulin sensitivity of triglyceride lipolysis in subcutaneous adipose (Stallknecht *et al.*, 2000). Also, data from transgenic animals have further supported the importance of adipose tissue in modulating whole-body insulin sensitivity (Shepherd *et al.*, 1993). Hence, the role of adipose tissue mediated by lipid metabolism in the physiological adaptation process to hot arid environments has been well documented. Furthermore, adipose tissue serves as an organ for storing food and a source of energy and water, contributing to animal survival during prolonged starvation and thirst. Moreso, fat oxidation produces metabolic water in water-deprived or exercised animals (Pond 2017). Strikingly, two other genes, *PLCH2* and *PEX10*, found within the unique Sudanese zebu selective sweep on chromosome 16, are involved in the adaptive metabolic strategy of lipid catabolism and oxidation. Functionally, the phospholipase C eta 2 (*PLCH2*) gene, also known as *PLCL4*, among other aliases, is suggested to participate in the lipid catabolic process for glucogenesis. Moreover, the hypermethylation of loci associated with *PLCH2* coinciding with fat reduction has also been reported in calory restricted humans

(Bouchard *et al.*, 2010). On the other hand, peroxisomes such as *PEX10* controls the composition of intracellular fatty acid content especially, the unsaturated fatty acid content, and are essential in the oxidation of fatty acids to produce water (Hong *et al.*, 2009, Fillmore *et al.*, 2011, Yan *et al.*, 2020).

In conclusion, the present study analyses the whole genome re-sequencing of six indigenous Sudanese zebu breeds to understand their genome diversity and the genomic footprints of their adaptations to environmental challenges in the African dryland area. We reported several gene-rich selective sweeps following three genomic scan approaches. The results highlight the importance of selection at gene-rich genome regions as a mechanism of adaptation to the complexity of environmental challenges. If under the genetic control of a similar regulatory mechanism, these gene-rich regions may contribute to the rapid adaptation of cattle to new challenging environments. Also, these regions could be enriched for superalleles, but this will require further investigation. Moreover, we identified a unique selective sweep on chromosome 16 overlapping multiple genes involved in insulin signalling, feed efficiency, glucose and fat metabolism, strongly suggesting their roles in adaptive metabolic strategies during feed and water deprivation and energy requirements. Our results align with previous studies that reported the relevance of these cellular processes in other species, including humans, enabling adaptation to similar conditions of the harsh desert environment.

Disclosures

The authors with this declare that there are no conflicts of interest, be it financial or otherwise

Author contributions

Author contributions: HM, OH, and AT designed the research; HM secured funding for the sampling and sequencing of the Group A Sudanese zebu cattle; MVB contributed the data on Gir cattle from Brazil; BS, HE and TM collected the Group A Sudanese samples; KM and BS collected and contributed the Group B Sudanese zebu samples; AT analysed the data; AT and OH interpreted the results; AT drafted the manuscript, AT and OH edited the manuscript; All authors revised and approved the manuscript's final version.

Grants

This work is part of the project “agricultural growth, capacity building for scientific preservation of livestock breeds in Sudan”. The project was supported by a Korea-Africa Economic Cooperation Trust Fund grant through the African Development Bank (Grant No: KOAFEC-TF-2013).

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 nor the UK Government.

Acknowledgements

The Group A Sudanese zebu breeds were preliminary analysed and presented in the University of Nottingham thesis of Dr Abdulfatai Tijjani. The authors acknowledge the Ministry of Finance and National Economic, Sudan, and the African Development Bank, country office (Khartoum, Sudan) to

Genomic signatures of desert adaptation

support the project's approval and implementation.

References

- Epstein, H. (1971). The origin of the domestic animals of Africa, Africana Publishing Corporation.
- Rahman, I. (2007). "Sudanese cattle resources and their productivity: a review." *Agricultural Reviews* 28(4): 305-308.
- Nahar, A. B. (2009). Sudan's fourth national report to the convention on biological diversity. Khartoum, Sudan: Higher Council for Environment and Natural Resources
- HCENR, (Higher Council for Environment and Natural Resources). (2014). Sudan's Fifth National Report to the Convention on Biological Diversity. Ministry of Environment, Natural Resources and Physical Development, Sudan. <https://www.cbd.int/doc/world/sd/sd-nr-05-en.pdf>.
- Impiglia, A. (2017). Study on small-scale family farming in the Near East and North Africa region. Focus country-Sudan. Study on small-scale family farming in the Near East and North Africa region. Focus country-Sudan.
- Musa, L., Peters, K. and Ahmed, M. (2006). "On farm characterization of Butana and Kenana cattle breed production systems in Sudan." *Livestock research for rural development* 18(12): 2006.
- Ahmed, S. K., Rahman, A. H., Hassan, M. A., Salih, E. M., Paone, M. and Cecchi, G. (2016). "An atlas of tsetse and bovine trypanosomosis in Sudan." *Parasites & vectors* 9(1): 194.
- Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R., Sherwood, E., Webster, M. T., Jiang, L., Ingman, M., Sharpe, T. and Ka, S. (2010). "Whole-genome resequencing reveals loci under selection during chicken domestication." *Nature* 464(7288): 587-591.
- Beaumont, M. A. (2005). Adaptation and speciation: what can F_{st} tell us?. *Trends in ecology & evolution*, 20(8), 435-440.
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., ... & Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), 913-918.
- Stothard, P., Choi, J.-W., Basu, U., Sumner-Thomson, J. M., Meng, Y., Liao, X. and Moore, S. S. (2011). "Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery." *BMC Genomics* 12(1): 559.
- Bahbahani, H., Tijjani, A., Mukasa, C., Wragg, D., Almathen, F., Nash, O., Akpa, G. N., Mbole-Kariuki, M., Malla, S. and Woolhouse, M. (2017). "Signatures of Selection for Environmental Adaptation and Zebu× Taurine Hybrid Fitness in East African Shorthorn Zebu." *Frontiers in genetics* 8: 68

- Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., Agaba, M., Kim, K., Kwak, W. and Sung, S. (2017). "The genome landscape of indigenous African cattle." *Genome biology* 18(1): 34.
- Weldenegodguad, M., Popov, R., Pokharel, K., Ammosov, I., Ming, Y., Ivanova, Z., & Kantanen, J. (2019). Whole-genome sequencing of three native cattle breeds originating from the northernmost cattle farming regions. *Frontiers in genetics*, 9, 728.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
- Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elsik, C. G., Tseng, E., ... & Medrano, J. F. (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. *GigaScience*, 9(3), giaa021.
- Li, H. and Durbin, R. (2009). "Fast and accurate short read alignment with burrows-wheeler transform." *Bioinformatics* 25.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009). "The sequence alignment/map format and SAMtools." *Bioinformatics* 25(16): 2078-2079.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S. and Daly, M. (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." *Genome research* 20(9): 1297-1303.
- Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., ... & Flicek, P. (2019). Ensembl 2019. *Nucleic acids research*, 47(D1), D745-D751.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T. and Sherry, S. T. (2011). "The variant call format and VCFtools." *Bioinformatics* 27(15): 2156-2158.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. and Ruden, D. M. (2012). "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(2): 80-92.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. and Daly, M. J. (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." *The American Journal of Human Genetics* 81(3): 559-575.
- Alexander, D. H., Novembre, J. and Lange, K. (2009). "Fast model-based estimation of ancestry in unrelated individuals." *Genome research* 19(9): 1655-1664.
- Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular ecology resources*, 17(1), 27-32.

Genomic signatures of desert adaptation

- Pickrell, J. K. and Pritchard, J. K. (2012). "Inference of population splits and mixtures from genome-wide allele frequency data." *PLoS genetics* 8(11): e1002967.
- Rubin, C.-J., Megens, H.-J., Barrio, A. M., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P. and Jørgensen, C. B. (2012). "Strong signatures of selection in the domestic pig genome." *Proceedings of the National Academy of Sciences* 109(48): 19529-19536.
- Axelsson, E., Ratnakumar, A., Arendt, M.-L., Maqbool, K., Webster, M. T., Perloski, M., Liberg, O., Arnemo, J. M., Hedhammar, Å. and Lindblad-Toh, K. (2013). "The genomic signature of dog domestication reveals adaptation to a starch-rich diet." *Nature* 495(7441): 360-364.
- Weir, B. S. and Cockerham, C. C. (1984). "Estimating F-statistics for the analysis of population structure." *Evolution* 38(6): 1358-1370
- Quinlan A.R., Hall I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841–842
- Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G. and Kasprzyk, A. (2009). "BioMart—biological queries made easy." *BMC Genomics* 10(1): 22.
- Gautier, M., Klassmann, A., & Vitalis, R. (2017). rehh 2.0: a reimplement of the R package rehh to detect positive selection from haplotype structure. *Molecular ecology resources*, 17(1), 78-90.
- Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, 81(5), 1084-1097.
- Dutta, P., Talenti, A., Young, R., Jayaraman, S., Callaby, R., Jadhav, S. K., ... & Prendergast, J. G. (2020). Whole genome analysis of water buffalo and global cattle breeds highlights convergent signatures of domestication. *Nature communications*, 11(1), 1-13.
- Kim, K., Kwon, T., Dessie, T., Yoo, D., Mwai, O. A., Jang, J., ... & Kim, H. (2020). The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nature Genetics*, 52(10), 1099-1110.
- Tijjani, A., Utsunomiya, Y. T., Ezekwe, A. G., Nashiru, O., & Hanotte, O. (2019). Genome sequence analysis reveals selection signatures in endangered trypanotolerant West African Muturu cattle. *Frontiers in genetics*, 10, 442.
- Porto-Neto, L. R., Reverter, A., Prayaga, K. C., Chan, E. K., Johnston, D. J., Hawken, R. J., ... & Barendse, W. (2014). The genetic architecture of climatic adaptation of tropical cattle. *PloS one*, 9(11), e113284.
- Naval-Sánchez, M., Porto-Neto, L. R., Cardoso, D. F., Hayes, B. J., Daetwyler, H. D., Kijas, J., & Reverter, A. (2020). Selection signatures in tropical cattle are enriched for promoter and coding regions and reveal missense mutations in the damage response gene HELB. *Genetics Selection Evolution*, 52, 1-19.

- Kissick, H. T., Dunn, L. K., Ghosh, S., Nechama, M., Kobzik, L., & Arredouani, M. S. (2014). The scavenger receptor MARCO modulates TLR-induced responses in dendritic cells. *PLoS one*, 9(8), e104148.
- Liao, X., Peng, F., Forni, S., McLaren, D., Plastow, G., Stothard, P. and Scherer, S. (2013). “Whole genome sequencing of Gir cattle for identifying polymorphisms and loci under selection 1.” *Genome* 56(10): 592-598.
- Rocha, J. L., Godinho, R., Brito, J. C., & Nielsen, R. (2021). Life in Deserts: The Genetic Basis of Mammalian Desert Adaptation. *Trends in Ecology & Evolution*.
- Nasteska, D., Fine, N. H., Ashford, F. B., Cuozzo, F., Vilorio, K., Smith, G., ... & Hodson, D. J. (2021). PDX1 LOW MAFA LOW β -cells contribute to islet function and insulin release. *Nature communications*, 12(1), 1-19.
- Cui, J., Wang, Z., Cheng, Q., Lin, R., Zhang, X. M., Leung, P. S., ... & Huang, J. D. (2011). Targeted inactivation of kinesin-1 in pancreatic β -cells in vivo leads to insulin secretory deficiency. *Diabetes*, 60(1), 320-330.
- Shim, H., Wu, C., Ramsamooj, S., Bosch, K. N., Chen, Z., Emerling, B. M., ... & Cantley, L. C. (2016). Deletion of the gene *Pip4k2c*, a novel phosphatidylinositol kinase, results in hyperactivation of the immune system. *Proceedings of the National Academy of Sciences*, 113(27), 7596-7601.
- Laskowski, D., Sjunnesson, Y., Humblot, P., Andersson, G., Gustafsson, H., & Båge, R. (2016). The functional role of insulin in fertility and embryonic development—What can we learn from the bovine model?. *Theriogenology*, 86(1), 457-464.
- Lagundžin, D., Hu, W. F., Law, H. C., Krieger, K. L., Qiao, F., Clement, E. J., ... & Woods, N. T. (2019). Delineating the role of FANCA in glucose-stimulated insulin secretion in β cells through its protein interactome. *PloS one*, 14(8), e0220568.
- Zou, L., Yan, S., Guan, X., Pan, Y., & Qu, X. (2013). Hypermethylation of the PRKCZ gene in type 2 diabetes mellitus. *Journal of diabetes research*, 2013.
- Ahmed, S. A. H., Ansari, S. A., Mensah-Brown, E. P., & Emerald, B. S. (2020). The role of DNA methylation in the pathogenesis of type 2 diabetes mellitus. *Clinical Epigenetics*, 12(1), 1-23.
- Diaz, M., Martel, N., Fitzsimmons, R. L., Eriksson, N. A., Cowin, G. J., Thomas, G. P., ... & Leong, G. M. (2012). Ski overexpression in skeletal muscle modulates genetic programs that control susceptibility to diet-induced obesity and insulin signaling. *Obesity*, 20(11), 2157-2167.
- Oleksyk, T. K., Smith, M. W., & O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1537), 185-205.
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., ... & Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), 913-918.

Genomic signatures of desert adaptation

- Naval-Sanchez, M., Nguyen, Q., McWilliam, S., Porto-Neto, L. R., Tellam, R., Vuocolo, T., ... & Kijas, J. (2018). Sheep genome functional annotation reveals proximal regulatory elements contributed to the evolution of modern breeds. *Nature communications*, 9(1), 1-13.
- Jay, P., Whibley, A., Frézal, L., de Cara, M. Á. R., Nowell, R. W., Mallet, J., ... & Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. *Current Biology*, 28(11), 1839-1845.
- Wu, H., Guang, X., Al-Fageeh, M. B., Cao, J., Pan, S., Zhou, H., ... & Wang, J. (2014). Camelid genomes reveal evolution and adaptation to desert environments. *Nature communications*, 5(1), 1-10.
- Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., & Rothschild, M. F. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*, 116(3), 255-264.
- Mwacharo, J. M., Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B. A., & Rothschild, M. F. (2017). Genomic footprints of dryland stress adaptation in Egyptian fat-tail sheep and their divergence from East African and western Asia cohorts. *Scientific reports*, 7(1), 1-10.
- Mirkena, T., Duguma, G., Haile, A., Tibbo, M., Okeyo, A. M., Wurzinger, M., & Sölkner, J. (2010). Genetics of adaptation in domestic farm animals: A review. *Livestock Science*, 132(1-3), 1-12.
- Gebreyohanes, M. G., & Assen, A. M. (2017). Adaptation mechanisms of camels (*Camelus dromedarius*) for desert environment: a review. *J. Vet. Sci. Technol*, 8(6), 1-5.
- Yurchenko, A. A., Daetwyler, H. D., Yudin, N., Schnabel, R. D., Vander Jagt, C. J., Soloshenko, V., ... & Larkin, D. M. (2018). Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Scientific reports*, 8(1), 1-16
- Aguiar, T. S., Torrecilha, R. B. P., Milanesi, M., Utsunomiya, A. T. H., Trigo, B. B., Tijjani, A., ... & Utsunomiya, Y. T. (2018). Association of copy number variation at intron 3 of HMGA2 with navel length in *Bos indicus*. *Frontiers in genetics*, 9, 627.
- Dixit, S. P., Bhatia, A. K., Ganguly, I., Singh, S., Dash, S., Sharma, A., ... & Jayakumar, S. (2021). Genome Analyses Revealed Genetic Admixture and Selection Signatures in *Bos indicus*.
- Kim, K. S., Thomsen, H., Bastiaansen, J., Nguyen, N. T., Dekkers, J., Plastow, G. S. and Rothschild, M. F. (2004). "Investigation of obesity candidate genes on porcine fat deposition quantitative trait loci regions." *Obesity* 12(12): 1981-1994.
- Weedon, M. N., Lettre, G., Freathy, R. M., Lindgren, C. M., Voight, B. F., Perry, J. R., ... & Zeggini, E. (2007). A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature genetics*, 39(10), 1245-1250.
- Boyko, A. R., Quignon, P., Li, L., Schoenebeck, J. J., Degenhardt, J. D., Lohmueller, K. E., Zhao, K., Brisbin, A., Parker, H. G. and Cargill, M. (2010). "A simple genetic architecture underlies morphological variation in dogs." *PLoS biology* 8(8): e1000451.

- 917
- 918 Fariello, M. I., Servin, B., Tosser-Klopp, G., Rupp, R., Moreno, C., San Cristobal, M., ... &
- 919 International Sheep Genomics Consortium. (2014). Selection signatures in worldwide sheep
- 920 populations. *PloS one*, 9(8), e103813.
- 921
- 922 Sevane, N., Dunner, S., Boado, A. and Cañon, J. (2017). “Polymorphisms in ten candidate genes are
- 923 associated with conformational and locomotive traits in Spanish Purebred horses.” *Journal of applied*
- 924 *genetics* 58(3): 355-361.
- 925
- 926 Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Bashir, S., Diallo, B., ... & Kim, H. (2017). The genome
- 927 landscape of indigenous African cattle. *Genome biology*, 18(1), 1-14.
- 928
- 929 Bahbahani, H., Tijjani, A., Mukasa, C., Wragg, D., Almathen, F., Nash, O., ... & Hanotte, O. (2017).
- 930 Signatures of selection for environmental adaptation and zebu× taurine hybrid fitness in East African
- 931 Shorthorn Zebu. *Frontiers in genetics*, 8, 68.
- 932
- 933 Kim, K., Kwon, T., Dessie, T., Yoo, D., Mwai, O. A., Jang, J., ... & Kim, H. (2020). The mosaic
- 934 genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nature*
- 935 *Genetics*, 52(10), 1099-1110.
- 936
- 937 Ahbara, A., Bahbahani, H., Almathen, F., Al Abri, M., Agoub, M. O., Abeba, A., ... & Mwacharo, J.
- 938 M. (2019). Genome-wide variation, candidate regions and genes associated with fat deposition and tail
- 939 morphology in Ethiopian indigenous sheep. *Frontiers in genetics*, 9, 699.
- 940
- 941 Nguyen, T. H., Turek, I., Meehan-Andrews, T., Zacharias, A., & Irving, H. (2020). Analysis of
- 942 interleukin-1 receptor associated kinase-3 (IRAK3) function in modulating expression of inflammatory
- 943 markers in cell culture models: A systematic review and meta-analysis. *Plos one*, 15(12), e0244570.
- 944
- 945 Dorrington, M. G., Roche, A. M., Chauvin, S. E., Tu, Z., Mossman, K. L., Weiser, J. N. and Bowdish,
- 946 D. M. (2013). “MARCO is required for TLR2-and Nod2-mediated responses to *Streptococcus*
- 947 *pneumoniae* and clearance of pneumococcal colonization in the murine nasopharynx.” *The Journal of*
- 948 *Immunology* 190(1): 250-258.
- 949
- 950 Benard, E. L., Roobol, S. J., Spaink, H. P. and Meijer, A. H. (2014). “Phagocytosis of mycobacteria
- 951 by zebrafish macrophages is dependent on the scavenger receptor Marco, a key control factor of pro-
- 952 inflammatory signalling.” *Developmental & Comparative Immunology* 47(2): 223-233.
- 953
- 954 Ma, M.-J., Wang, H.-B., Li, H., Yang, J.-H., Yan, Y., Xie, L.-P., Qi, Y.-C., Li, J.-L., Chen, M.-J. and
- 955 Liu, W. (2011). “Genetic variants in MARCO are associated with the susceptibility to pulmonary
- 956 tuberculosis in Chinese Han population.” *PLoS One* 6(8): e24069.
- 957
- 958 Bowdish, D. M., Sakamoto, K., Lack, N. A., Hill, P. C., Sirugo, G., Newport, M. J., Gordon, S., Hill,
- 959 A. V. and Vannberg, F. O. (2013). “Genetic variants of MARCO are associated with susceptibility to
- 960 pulmonary tuberculosis in a Gambian population.” *BMC medical genetics* 14(1): 47.
- 961
- 962 Thuong, N. T. T., Tram, T. T. B., Dinh, T. D., Thai, P. V. K., Heemskerk, D., Bang, N. D., Chau, T.
- 963 T. H., Russell, D. G., Thwaites, G. E. and Hawn, T. R. (2016). “MARCO variants are associated with
- 964 phagocytosis, pulmonary tuberculosis susceptibility and Beijing lineage.” *Genes and immunity* 17(7):
- 965 419.

Genomic signatures of desert adaptation

- Ocepek, M., Pate, M., Žolnir-Dovč, M. and Poljak, M. (2005). "Transmission of *Mycobacterium tuberculosis* from human to cattle." *Journal of clinical microbiology* 43(7): 3555-3557.
- Thoen, C. O., Steele, J. H. and Gilsdorf, M. J. (2008). *Mycobacterium bovis* infection in animals and humans, John Wiley & Sons.
- Suliman, M. and Hamid, M. (2002). "Identification of acid fast bacteria from caseous lesions in cattle in Sudan." *Zoonoses and Public Health* 49(9): 415-418.
- Ibrahim, S., Abubakar, U., Danbirni, S., Usman, A., Ballah, F., Kudi, A., Lawson, L., Abdulrazak, H. and Abdulkadir, I. (2016). "Molecular identification of *Mycobacterium tuberculosis* transmission between cattle and man: a case report." *J Microbiol Exp* 3(Suppl 3): 00091.
- Jonsson, N. N. (2006). The productivity effects of cattle tick (*Boophilus microplus*) infestation on cattle, with particular reference to *Bos indicus* cattle and their crosses. *Veterinary parasitology*, 137(1-2), 1-10.
- Uddin, M. E., & Kebeab, E. (2020). Impact of Food and Climate Change on Pastoral Industries. *Frontiers in Sustainable Food Systems*, 4, 200.
- Cartee, G., Kietzke, E. and Briggs-Tung, C. (1994). "Adaptation of muscle glucose transport with caloric restriction in adult, middle-aged, and old rats." *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 266(5): R1443-R1447.
- Röpke, R., Schams, D., Schwarz, F. and Kirchgeßner, M. (1994). "Growth-related hormones in plasma of bulls, steers and heifers given food with two different energy levels." *Animal Science* 59(3): 367-377.
- Sternbauer, K. and Luthman, J. (2002). "Insulin sensitivity of heifers on different diets." *Acta Veterinaria Scandinavica* 43(2): 107.
- Barretero-Hernandez, R., Galyean, M. and Vizcarra, J. (2010). "The effect of feed restriction on plasma ghrelin, growth hormone, insulin, and glucose tolerance in pigs." *The Professional Animal Scientist* 26(1): 26-34.
- Leong, G. M., Kee, A. J., Millard, S. M., Martel, N., Eriksson, N., Turner, N., ... & Muscat, G. E. O. (2010). The Ski proto-oncogene regulates body composition and suppresses lipogenesis. *International journal of obesity*, 34(3), 524-536.
- Fitzsimons, C., Kenny, D., Waters, S., Earley, B. and McGee, M. (2014). "Effects of phenotypic residual feed intake on response to a glucose tolerance test and gene expression in the insulin signaling pathway in longissimus dorsi in beef cattle." *Journal of animal science* 92(10): 4616-4631.
- Kriebel, J., Herder, C., Rathmann, W., Wahl, S., Kunze, S., Molnos, S., ... & Grallert, H. (2016). Association between DNA methylation in whole blood and measures of glucose metabolism: KORA F4 Study. *PloS one*, 11(3), e0152314.

- Keogh, K., Kenny, D. A., Kelly, A. K. and Waters, S. M. (2015). “Insulin secretion and signaling in response to dietary restriction and subsequent re-alimentation in cattle.” *Physiological genomics* 47(8): 344-354.
- Keogh, K., Kenny, D. A., Kelly, A. K., & Waters, S. M. (2015). Insulin secretion and signaling in response to dietary restriction and subsequent re-alimentation in cattle. *Physiological genomics*, 47(8), 344-354.
- Scheepers, A., Joost, H. G., & Schurmann, A. (2004). The glucose transporter families SGLT and GLUT: molecular basis of normal and aberrant function. *Journal of parenteral and enteral nutrition*, 28(5), 364-371.
- Stallknecht, B., Larsen, J. J., Mikines, K. J., Simonsen, L., Bülow, J., & Galbo, H. (2000). Effect of training on insulin sensitivity of glucose uptake and lipolysis in human adipose tissue. *American Journal of Physiology-Endocrinology And Metabolism*, 279(2), E376-E385.
- Shepherd, P. R., Gnudi, L., Tozzo, E., Yang, H., Leach, F., & Kahn, B. B. (1993). Adipose cell hyperplasia and enhanced glucose disposal in transgenic mice overexpressing GLUT4 selectively in adipose tissue. *Journal of Biological Chemistry*, 268(30), 22243-22246.
- Pond, C. M. (2017). The evolution of mammalian adipose tissues. In *Adipose tissue biology* (pp. 1-59). Springer, Cham.
- Bouchard, L., Rabasa-Lhoret, R., Faraj, M., Lavoie, M. È., Mill, J., Pèrusse, L., & Vohl, M. C. (2010). Differential epigenomic and transcriptomic responses in subcutaneous adipose tissue between low and high responders to caloric restriction. *The American journal of clinical nutrition*, 91(2), 309-320
- Hong, S. P., Sharpe, P. L., Xue, Z., Yadav, N. S., & Zhu, Q. Q. (2009). U.S. Patent Application No. 12/244,950.
- Fillmore, N., Alrob, O. A., & Lopaschuk, G. D. (2011). Fatty Acid beta-Oxidation. *AOCS Lipid library*, doi, 10.
- Yan, F. X., Dong, G. R., Qiang, S., Niu, Y. J., Hu, C. Y., & Meng, Y. H. (2020). Overexpression of $\Delta^{12,\Delta^{15}}$ -Desaturases for Enhanced Lipids Synthesis in *Yarrowia lipolytica*. *Frontiers in microbiology*, 11, 289.

Genomic signatures of desert adaptation

Supplementary Figures

Figure S1 - The distributions of Z-transformed Hp scores across the bovine autosome in three zebu breeds (non-Sudanese); Kenya Boran (A), Ethiopian Ogaden (B) and Gir from Brazil (C). The blue line indicates the ZH_p threshold value (lowest 0.5%) for selecting outlier windows (candidate regions under positive selection).

Genomic signatures of desert adaptation

Figure S2 – The distribution of XP-EHH scores along Bovine chromosome 16 following the comparison of Sudanese zebu population with other zebu and taurine cattle populations. Sudanese Fulani (A), other zebu breeds (East African and Asian zebu) (B), and Taurine breeds (West African and Euroasia and West Europe breeds) (C). The red dash blue line indicates the significant XP-EHH threshold of >4 ($p_value < 0.001$) for selecting outlier SNPs (SNPs targeted by positive selection).

Genomic signatures of desert adaptation

Figure S3A – The decay of extended haplotype homozygosity around six significant SNPs in the Sudanese zebu population

Figure S3B – The decay of extended haplotype homozygosity around six significant SNPs in Sudanese Fulani

Figure S3C – The decay of extended haplotype homozygosity around six significant SNPs in other zebu population