Genome-wide scan for selection signatures reveals novel insights into the adaptive capacity in local North African cattle

Slim Ben-Jemaa[1*], Salvatore Mastrangelo[2], Seung-Hwan Lee[3], Jun Heon Lee[3], Mekki Boussaha[4]

[1]Laboratoire des Productions Animales et Fourragères, Institut National de la Recherche Agronomique de Tunisie, Université de Carthage, 2049 Ariana, Tunisia.

[2]Dipartimento Scienze Agrarie, Alimentari e Forestali, University of Palermo, 90128 Palermo, Italy.

[3]Division of Animal and Dairy Science, Chungnam National University, Daejeon, Korea

[4]INRAE, AgroParisTech, GABI, Université Paris Saclay, 78350, Jouy-en-Josas, France.

[*]Corresponding author*:* Slim Ben-Jemaa

Email: benjemaaslim@gmail.com

Tel: +216-53498579

Abstract

Natural-driven selection is supposed to have left detectable signatures on the genome of North African cattle which are often characterized by the fixation of genetic variants associated with traits under selection pressure and/or an outstanding genetic differentiation with other populations at particular loci. Here, we investigate the population genetic structure and we provide a first outline of potential selection signatures in North African cattle using SNP genotyping data. After comparing our data to African, European and indicine cattle populations, we identified 34 genomic regions using three extended haplotype homozygosity statistics and 88 outlier markers based on Bayescan test. The 14 outlier windows detected by at least two approaches, harboured genes (e.g. *GH1, ACE, ASIC3, HSPH1, MVD, BCL2*, *HIGD2A*, *CBFA2T3*) that may be involved in physiological adaptations required to cope with environmental stressors that are typical of the North African area such as infectious diseases, extended drought periods, scarce food supply, oxygen scarcity in the mountainous areas and high-intensity solar radiation. Our data also point to candidate genes involved in transcriptional regulation suggesting that regulatory elements had also a prominent role in North African cattle response to environmental constraints. Our study yields novel insights into the unique adaptive capacity in these endangered populations emphasizing the need for the use of whole genome sequence data to gain a better understanding of the underlying molecular mechanisms.

Introduction

Taurine cattle were first introduced to Africa through Egypt from the Fertile Crescent ~6500 years BP[1] and dispersed into North Africa where they have undergone hybridization with local wild aurochs[2]. The geographic proximity of North Africa to Europe makes it a likely contact zone between the two continents. Several genetic studies reported an old presence of African cattle ancestry in the genomes of Iberian cattle[2,3] and a European ancestry in local Maghreb cattle[4–6]. Nomad pastoralism and tribal migrations prevented the division of North African cattle populations into clearly defined breed groups. Present-day indigenous cattle in Morocco, Algeria Tunisia and Libya belong to the Brown Atlas cattle. These are small-sized, sturdy, fairly compact animals with fine limbs, a short head and a straight to slightly concave profile. In these countries, Brown Atlas cattle populations, predominantly pasture-fed, are raised in a Mediterranean climate characterized by a winter rainfall and a hot dry summer during which live weight losses in adult cows can reach 20%[7]. In Egypt, indigenous cattle are medium sized, long-bodied animals, lean of musculature and lightly boned. They are raised either in desert or semi-desert regions characterized by a very arid Mediterranean climate and negligible rainfall. A number of ecotypes are recognised based on their geographical distribution. For instance, In Lower Egypt there are two local cattle populations, the Damietta is typically found in coastal sites and the Baladi or Baheri is widespread inland in the delta[8]. Overall, North African indigenous cattle are resistant to many of the diseases and parasites to which imported European cattle are susceptible[7] resulting from a local environment-driven selection that occurred over hundreds of years. Adaptation to local conditions is expected to leave distinct signatures in the genome known as a "selective sweeps" owing to a rapid increase in the frequency of the desirable alleles or in the frequency of neutral markers in linkage disequilibrium with the favorable alleles[9]. Reports on signatures of selection focusing exclusively on North African cattle have never been reported before.

The emergence of high-throughput single nucleotide polymorphism (SNP) genotyping and whole genome sequencing facilities coupled with the development of new genomic methodologies have

enabled the screening of a large part of the genome to detect signatures of selection in livestock and domestic populations[10–14]. All these studies have used comparison of genomic patterns of SNPs variability between local and exotic breeds to identify genomic regions and genes that have undergone selective sweeps.

The main goal of this study was to investigate population structure and candidate positive selection signatures in North African cattle using genotype data from the Illumina BovineSNP50 BeadChip with comparisons against four European breeds, three African and two indicine populations. We applied four genome scan approaches to identify genomic regions putatively under selection: the first three methods are extended haplotype homozygosity (EHH)-derived statistics (*iHS*, *Rsb* and *XP-EHH*) and are based on the decay of haplotype homozygosity as a function of recombination distance. The fourth approach is a Bayesian method based on the differentiation of allele frequencies among populations.

Results

Population structure analysis among all cattle populations

We used Principal Component Analysis (PCA) to contextualize the genetic variation of North African cattle populations (Fig. 1). The first two principal components accounted for 5.86 % (PC1) and 3.75 % (PC2) of the total genetic variation. The global organization of the genetic diversity of the populations of the study might be described as a triangle with apexes corresponding to North European breeds (Angus (ANG) and Holstein (HOL)), African taurines (NDA, ND1 and ND2) and indicine populations (NEL and GIR). PCA results show that the Tunisian Brune de l'Atlas (TUNIND) and the Algerian populations (Guelmoise (GUE) and Cheurfa (CHE)) are closer to each other than to the Moroccan (Oulmes Zaer (OUL) and Tidili (TID)) and the Egyptian (Baladi (BAL)) populations. Furthermore, these results distinguished Biskra (BIS) and Chelifienne (CHF) from the other North African populations. The former was positioned near European breeds with several BIS individuals clustering along with Montbeliarde (MON) while CHF individuals showed a higher

dispersion around their center of gravity (with several individuals positioned near MON) indicating a high genetic heterogeneity.

Breed assignment to clusters using ADMIXTURE provided further insight into the genetic structure of North African populations. Fig. 2 shows the results obtained for K values 2, 3, 5, 7, 10, 12 and 17. K=12 showed the lowest cross-validation error (Supplementary Fig. S1). As expected, admixture analysis at K = 2 separated taurine from indicine cattle reflecting their ancient divergence. Among North African populations, BAL has the highest indicine ancestry (on average, 33.6% with a minimum of 19.9% and maximum of 37.8%). The K = 3 model separated African from European breeds. All North African populations except BAL carry two main European and African ancestries. In agreement with PCA results, BIS shows the largest amount of European ancestry with a minimum of 62.37% and a maximum of 90.29% while the Moroccan TID has the largest amount of African ancestry with a minimum of 55.13% and a maximum of 69.95%. For its part, BAL possesses a significant amount of indicine ancestry with a minimum of 16.44% and a maximum of 30.39%. At K=5, the three European breeds (ANG, HOL and Jersey (JER)), formed three different clusters. All North African populations had on average 20% (with a minimum of 10% in BAL and a maximum of 26.76% in BIS) and 18.5% (with a minimum of 10.55% in BAL and a maximum of 45.88% in BIS) of JER and HOL ancestries, respectively. At K=7, all North African populations except BIS and a few CHF individuals can be seen as distinct from the other breeds with a major "North African" component ranging, on average, from 48.7% for BAL and CHF to 78.86% for TID. It is worth noting that BIS displayed a substantial level of MON introgression (on average, 32.4%) while no African ancestry was detected within this breed (Fig. 2). At K=12, BAL separated from the other North African populations while this happened for OUL when K was set to 17.

Details of the level of pairwise genetic differentiation are reported in Supplementary Table S1. Most of North African populations showed low differentiation levels. The lowest F_{ST} values are found between CHE and GUE (F_{ST} ~0), CHE and TUNIND (F_{ST} = 0.002) and between GUE and

TUNIND (F_{ST} = 0.003). Likewise, low genetic differentiation is observed between TID on one hand, GUE, CHE and TUNIND, on the other hand (F_{ST} TID/GUE= 0.016, F_{ST} TID/CHE=0.015 and F_{ST} TID/TUNIND=0.015) while a higher F_{ST} is observed between these three breeds and BAL (0.042, 0.042 and 0.044 for BAL/CHE, BAL/GUE and BAL/TUNIND, respectively).

We used the TreeMix software to model both population splits and gene flow between the 17 cattle populations. When no migration events were fit (Supplementary Fig. S2, residuals presented in Supplementary Fig. S3), the eight North African populations were positioned on different locations on the tree. BAL and TID were the closest to indicine and African populations, respectively while BIS was in clade with the European breeds. We then sequentially added migration events to the tree until the proportion of the variance in relatedness between populations explained by the model began to asymptote. This happened when 14 migration edges were fit (where 99.92% of the variance in ancestry between populations was explained by the model (Supplementary Fig. S4)). The phylogenetic network structure presented in Figure 3 highlights the known African taurine introgression into North African populations. In addition, CHF received a migration edge that originated from GUE while BIS and the Tunisian Brune de l'Atlas individuals showed a low-level of introgression from Holstein (HOL) and indicine breeds, respectively.

Candidate genome regions putatively under selection in North African cattle

In order to perform an accurate search for signatures of selection in North African cattle, we selected the breeds that are most representative of the ancestral North African populations i.e those with a major "North African" component. This was done based on population structure results and led to the exclusion of BIS (because of the low portion of its North African ancestry) and CHF (because of its high inter-individual genomic heterogeneity) (Figs. 1, 2 and 3). We also removed a total of 1,475 SNPs because of uncertainty in the identification of their ancestral state (see methods section).

Rsb and Cross-population Extended Haplotype Homozygosity (*XP-EHH)* statistics were computed at each SNP for each of the three comparisons (African (AFT)/North African, European

(EUT)/North African, indicine (IND)/North African). Haplotypes estimated in each population were pooled, for each autosome, according to their group of origin. In total, 112, 366 and 96 haplotypes were considered as representative of African, European, and indicine ancestries, respectively.

Rsb detected 381, 391 and 158 SNPs putatively under selection for AFT/North AFT, EUT/North AFT and /North AFT comparisons, respectively (Figs. 4a, 4b and 4c, respectively). These markers defined 12, 11 and 4 candidate regions for the comparisons between North AFT and AFT, North AFT and EUT and North AFT and IND, respectively (Figure 4, Table 1). *XP-EHH* yielded fewer outlier SNPs than analyses based on the *Rsb* approach: 254, 215 and 110 SNPs putatively under selection for AFT/North AFT, EUT/North AFT and IND/North AFT comparisons, respectively (Figs. 5a, 5b and 5c, respectively). These outliers defined 8, 7 and 3 selective sweeps for the comparisons between North AFT and AFT, North AFT and EUT and North AFT and IND, respectively (Table 1). Among these, six, four and two regions were also identified with Rsb tests for AFT/North AFT, EUT/North AFT and IND/North AFT comparisons, respectively (Table 1). These regions are located on chromosomes (BTA) 01 (at position: 17,700,000-19,640,000 bp), BTA04 (at positions :76,470,000-78,910,000 bp and 113,060,000-114,940,000 bp), BTA06 (at position : 46,780,000-50,050,000 bp) and BTA24 (at positions :18,030,000-20,020,000 bp and 59,780,000-61,840,000 bp) for the AFT/North AFT comparison, on BTA01 (at position: 82,820,000-84,810,000 bp), BTA07 (at position :41,100,000-43,620,000 bp), BTA19 (at position: 47,120,000-49,070,000 bp) and BTA21 (at position :14,830,000-16,650,000 bp) for the EUT/North AFT comparison and on BTA12 (at position: 28,540,000-30,300,000 bp), BTA18 (at position :11,580,000-14,300,000 bp) for the IND/North AFT comparison. The intra-population *iHS* analysis revealed a total of 4 significant regions (p$iHS \geq 3$) distributed on BTA 03, 08, 19 and 22 (Fig. 5d, Table 1) which had an average size of 1.64 Mb (ranging from 1.55 to 1.85 Mb). Among these, two candidate regions (BTA19: 47,390,000-48,980,000 bp and BTA22: 4,790,000-6,620,000 bp) overlapped with outlier windows detected by the EUT/North African (both *Rsb* and *XP-EHH* tests) and IND/North African (*XP-EHH* test) comparisons, respectively. Overall, the 13 candidate

genomic regions jointly identified by *Rsb, XP-EHH* and *iHS* statistics overlap with Quantitative Trait Loci (QTL) associated with traits for milk and meat composition, fertility and sexual precociousness, disease susceptibility (tuberculosis and respiratory diseases), stature and growth (Supplementary Table S2). Also, the 13 aforementioned genomic regions co-localized with 181 previously described structural variants most of which (161 out of 181) are copy number variations (CNV) (Supplementary Table S3). In total, 76 genes are located in CNV regions (Supplementary Table S4).

Bayesian F_{ST} method

We used the BayeScan program to identify putative genomic regions under selection in North African cattle. A total of 54 and 34 outlier SNPs were detected for F_{ST} AFT/North AFT and F_{ST} EUT/North AFT, respectively (Supplementary Fig. S5, Supplementary Tables S5 and S6). Among these 88 SNPs, only five markers were located within or close to candidate regions detected by an EHH-based metric (Supplementary Tables S5 and S6). No significant SNPs were identified with the F_{ST} IND/North AFT test.

Identification and functional annotation of the genes within the candidate regions

Outlier windows from *iHS*, *Rsb* and *XP-EHH* tests include 70, 570 and 318 known genes, respectively (Table 1). Genes identified with *Rsb* (*XP-EHH*) are distributed as follows : 166 (116), 285 (155) and 144 (51) for AFT/North African, EUT/North African, IND/North African comparisons, respectively (Table 1). Thirty six and eight genes were common to both *iHS* and EUT/North African (either *Rsb* or *XP-EHH*) and *XP-EHH* IND/North African comparisons, respectively. Similarly, 112, 180 and 56 genes were jointly identified by *Rsb* and *XP-EHH* for each of the AFT/North African, EUT/North African, IND/North African comparisons, respectively of which 76, 124 and 42, respectively, could be mapped by DAVID Bioinformatics resources (https://david.ncifcrf.gov/). Among the 21 identified functional term clusters (Supplementary Tables S7, S8 and S9), three are significantly enriched (Benjamini-corrected p-value < 0.05) relative to the whole bovine genome: one cluster identified in the *Rsb* AFT/North AFT comparison

(Supplementary Table S7) associated with GTP-binding activity (n = 9, Benjamini-corrected p-value = 0.0139) and two clusters identified in the *Rsb* EUT/North AFT comparison (Supplementary Table S8) associated with sensory perception of smell (n = 18, Benjamini-corrected p-value = 2.40×10^{-12}) and serine-type endopeptidase activity (n = 7, Benjamini-corrected p-value = 0.0283).

Discussion

The main purpose of the present study is to unravel signatures of positive selection in North African cattle. Because we used several breeds with diverse population structure, the main challenge in our study was to minimize the rate of false-positive signals that can arise, inter alia, owing to the confounding effects of population demographics[15]. Assuming that populations with similar structure have undergone similar evolutionary processes, in our selection signature detection analyses, we retained only North African populations showing a high degree of within population genetic homogeneity and a large portion of North African ancestry. In agreement with previous studies[6] our genome analyses are consistently and strongly in the direction of a substantial and recent contribution of European breeds to the genomes of BIS and CHF (Figs. 1 and 2). Furthermore, in the admixture models in which K = 12, the individuals sampled from these two breeds showed a high degree of within population genetic heterogeneity. Therefore, BIS and CHF were discarded from the subsequent selection signature analyses.

Our results corroborate previous reports[16] suggesting that BAL resulted from a three-way admixture between breeds representative of European, African and indicine cattle. The presence of an indicine content within the genome of BAL is consistent with a wave of indicine introduction during the rinderpest epidemic of the 19th century[1,17]. Our results indicate that all North African populations share ancestry with Jersey cattle which supports previous whole genome sequencing analyses reporting a common distinct patriline of Jersey bulls with African cattle[18]. Overall, our findings indicate that modern North African cattle can be classified into 3 subgroups. The first one is the "Brune de l'Atlas" population which possesses two main African and European ancestries. This

subgroup includes the Moroccan TID, the Algerian GUE and CHE and the Tunisian TUNIND. The second subgroup consists of the Egyptian local cattle which possesses an additional large portion of indicine ancestry (at the expense of European ancestry). The third subgroup, represented by CHF and BIS, includes European-derived breeds. The phylogenetic network inferred by TreeMix corroborate these findings in that CHF and especially BIS are in clade with the European breeds while CHE, TID, TUNIND and GUE share the same branch and are much closer to African populations.

In this paper, we present the first genome-wide scan of putative selective sweeps in North African cattle by combining four different statistical methods based either on the decay of haplotype homozygosity as a function of recombination distance or on allele frequency differentiation among populations. In total, we highlight the presence of 34 different genomic regions putatively under selection using the first type of approaches (*iHS, Rsb* and *XP-EHH*) and 88 outlier SNPs using Bayescan. Consistently with previous observations[19], we observe little overlap between results obtained from each of the two types of approaches. Given that Bayescan assumes that the gene frequencies under any neutrally structured population model can be approximated by a multinomial Dirichlet distribution[20] which would not be appropriate in a hierarchical population structure[21] (as is the case for our North African sample), the 88 identified SNPs should be considered cautiously. Instead, we believe that the three EHH-based methods, which inter alia, can detect a wider range of selection scenarios[22], are more suitable to our study design. These statistics take advantage of the reduction in haplotype diversity in the neighbourhood of a beneficial mutation due to a "hitch-hiking" effect. They measure the extended haplotype homozygosity which is defined as the probability of identity by descent for two randomly chosen haplotypes carrying a core haplotype of interest in an interval around a given locus, given that they have the same allele at the locus[23]. Unlike *Rsb* and *XP-EHH,* the *iHS* test has low power in identifying fixed sweeps because it requires the ancestral allele to be still segregating in the population[24]. Here, we identified a higher number of outlier windows using *Rsb* and *XP-EHH* compared to the *iHS* approach which might suggest, at

first glance, that most of the candidate regions identified here have undergone a positive selection resulting in the (near) fixation of the favoured alleles across the populations. However, we believe that the low number of candidate regions identified by the *iHS* test is actually due to the fact that this approach searches for loci where a given high-frequency haplotype is much longer relative to all other haplotypes, yet in a soft sweep several long haplotypes will be present at the adaptive locus and thus not one haplotype will typically be much longer than all others[25]. Our hypothesis assumes that the majority of sweeps detected here are soft which is likely to be the case. Soft sweeps were shown to be widespread and account for the vast majority of recent environmental adaptation in several species such as Humans[24]. A common constraint of selection signature detection methods is the detection of false positives. One efficient way to reduce their number is to retain as outliers, those genomic regions detected by distinct methods[26]. Among the 34 genomic regions identified by the three EHH-based methods, 13 were detected by at least two tests. In addition, a fourteenth region (BTA07: 36,720,000-38,670,000 bp) identified by the *Rsb* EUT/North AFT comparison included an outlier SNP detected by Bayescan. We particularly focused on genes located within these 14 genomic regions. In agreement with previous findings[27,28], we observed that the four candidate regions jointly identified by the *Rsb* and *XP-EHH* tests in the EUT/North African comparison were significantly enriched for genes involved in olfactory receptor activity (17 genes) which might reflect the fact that selection has been relaxed around these genes in European breeds which are often raised in abundant food supply conditions. Two genes (*OR2W3* and *OR2L13*) coincided with CNVs previously reported in cattle (Supplementary Table S4). Olfactory receptor genes are duplicated within the bovine genome[27] and CNVs encompassing these genes were found to be associated with population-specific differences in smell in most mammalian species[29].

Many of our candidate regions harboured genes implicated in the adaptive immune response against microbial pathogens. For instance, the clearest sweep signal in the EUT/North AFT comparison detected on BTA07 (between positions: 41.10 and 43.62 Mb) with 12 SNPs (out of 21) exceeding the significance threshold, harboured 57 known genes amongst which six (*AZU1*, *ELANE*, *GZMM*,

PRSS57, PRTN3, CFD) belong to the S1A family of peptidases, a superfamily of proteolytic enzymes with a wide variety of biological functions in parasite infection[30]. Likewise, the fourteenth aforementioned region on the BTA07 (at position : 36,720,000-38,670,000 bp) also harboured several genes directly associated with adaptive and antiviral immune responses (*COMMD10*, *FAF2, TSPAN17)*. Similarly, another relevant selection signature on the BTA19 jointly detected by *iHS*, *Rsb* and *XP-EHH* EUT/North African harboured several genes which are involved in immune response: *CD79B*, *MILR1, PECAM1, MAP3K3* and *TCAM1*. The last two genes mediate NF-kappa-B activity which show evidence of positive selection in the African N'Dama cattle to alter in functions to effectively regulate the infection of cattle trypanosome[31]. Consistently, we also observed that outlier windows from AFT/North African and IND/North African comparisons included many genes associated with immune response and host defence such as *TNFRSF11A*, *IRF8*, *MYO1G, TGFBR2* and several GTPases of immunity-associated protein (GIMAP) genes (GIMAP4, GIMAP5 and GIMAP7). Several of these genes (GIMAP4, GIMAP5, GIMAP7, *IRF8*) coincided with CNVs reported in cattle (Supplementary Table S4). A major phenotype of North African cattle populations is their resistance to parasitic diseases such as theileriosis, babesiosis and anaplasmosis[32] which are highly prevalent in North Africa[33]. We suggest that the aforementioned genes have been under evolutionary pressure in North African cattle and that some of them may have experienced enhanced fixation of duplicates resulting from selection for increased dosage to effectively regulate the innate and acquired immune response to parasitic diseases. A previous study[34] conducted on Brazilian *Bos indicus* cattle, similarly reported that CNVs are important modulators of immune gene expression. Our results have also revealed a series of other genes involved in the regulation of blood pressure and heart contraction (*ACE*, *ACE3*, *COX4I1*, *NOS3, CXADR*), blood vessel development and morphogenesis (*CCM2*, *FOXC2*, *FOXF1*, *MAP3K3*, *TGFBR2*). These genes are expected to be involved in adaptation to extreme temperatures prevailing in several Northern African areas and/or to chronic hypoxia in the Atlas mountain ranges where the altitude varies between 900 and 4,000 meters[7]. Our hypothesis is consistent with the

presence of three hypoxia-related genes (*BCL2*, *HIGD2A* and *CBFA2T3*) and three other genes involved in response to heat (*ASIC3*, *HSPH1* and *MVD)* in the relevant candidate regions (Table 1). It is also interesting to note that the strong selection signal on BTA19 harboured a well-known gene, *GH1,* linked to response to nutrient levels (GO: 0031667), positive regulation of lactation (GO:1903489) and triglyceride biosynthetic process (GO:0010867) and was previously reported as being a candidate gene for dairy production traits in Braunvieh cattle[15]. Importantly, it has been suggested that elevated *GH1* gene expression may constitute an adaptive response to the effects of scarce food supply in a sample of 163 human individuals from Benin[35]. We therefore suggest that this gene is particularly under positive selection across North African cattle populations as a consequence of important seasonal fluctuations in food availability characterizing the whole region.

Six out of the 14 relevant candidate regions identified in this study, harboured fewer than 15 known protein coding genes (Table 1). Many of these genes have also been reported in cattle and other species. For instance, the outlier window on BTA01 (at position: 17,700,000-19,640,000 bp), contained 6 protein coding genes including *TMPRSS15* and *CHODL*, two genes that were reported to be under selection in the Iraqi indigenous cattle[13]. Similarly, the candidate region on the BTA24 (at position: 59,780,000-61,840,000 bp), harboured *RNF152* gene which positively regulates Toll- like receptors (TLRs) which are important pattern recognition receptors that are critical for the defence against invading pathogens[36]. *RNF152* gene was reported to be involved in local adaptations in the Ainu, a hunter-gatherer population of northern Japan[37]. Another relevant candidate region on BTA21 (at position : 14,830,000-16,650,000 bp) harboured four protein coding genes : *SLCO3A1, SV2B, AKAP13* and *KLHL25.* The latter two genes were shown to be under positive selection in Creole cattle breeds[38] while *SLCO3A1* showed evidence of differential co-expression that has been shown to be under positive selection in the desert-adapted cactus mouse, *Peromyscus eremicus*[39]. *SV2B* gene was among major genes enriched for the extracellular matrix (ECM) around the hair follicle in Changthangi goats[40]. ECM is considered important for regulating the structure, metabolism and signaling of dermal papilla cells which play key roles in hair follicle

morphogenesis and regeneration[41]. Another candidate region on the BTA22 (at position: 4,790,000-6,620,000 bp) harboured four genes: *GADL1, TGFBR2*, *STT3B* and *OSBPL10. GADL1* gene is one of the genes involved in adaptive evolution of *Anolis carolinensis* introduced into the Ogasawara archipelago[42]. Gadl1$^{-/-}$ mice exhibited decreased anxiety, increased levels of oxidative stress markers, alterations in energy and lipid metabolism, and age-related changes[43]. *STT3B* is a catalytic subunit of hetero oligomeric oligosaccharyltransferase (OST), which is important for asparagine linked glycosylation. In mammals and plants, OSTs exhibit distinct levels of enzymatic efficiency or different responses to stressors[44]. *OSBPL10* gene confers African-ancestry protection against dengue haemorrhagic fever in admixed Cubans[45]. A further result is that the 14 outlier windows identified by at least two approaches included myriad of genes involved in transcriptional regulation (*AEBP1, ARID3A, BANP, CBFA2T3, DDX5, FTSJ3, GLI3, MIER2, POLR2E, POLRMT, FOXC2, FOXF1, FOXL1, SMARCD2, SMARCD3, TNFRSF11A, YEATS2, BPTF, CDK5, ...)* as well as many non-coding RNAs including 13 small nucleolar RNAs (snoRNAs), 10 microRNAs (miRNAs), 12 small nuclear RNAs (snRNA) and 15 long noncoding RNAs (lncRNAs). In addition, many of the aforementioned genes (*BANP, CBFA2T3, GLI3, POLR2E, POLRMT, FOXC2, FOXF1, FOXL1*) co-localize with known cattle CNVs. It is worth noting that CNVs encompassing a gene encoding a transcription factor has a greater phenotypic impact because it can affect both the coding sequence of the gene itself as well as the expression of downstream targets of that gene. From a selective standpoint, these findings suggest that natural selection has shaped North African cattle genome not only through variation in coding sequence but also through extensive regulation of gene expression occurring both at the transcriptional and post-transcriptional level. Lending further support to this hypothesis, the relevant candidate region on BTA24 (at position: 59,780,000-61,840,000 bp) harbours a single gene, *CELF4,* coding for an RNA-binding protein mainly expressed in central nervous system that regulates the expression of many genes co-transcriptionally or post-transcriptionally via interactions with mRNA[46]. Celf4-deficient mice have additional neurological abnormalities including hyperactivity and hyperphagia-associated obesity[47]. Similarly,

the most relevant selection signal in the AFT/North AFT comparison (BTA06 at position: 46,650,000 - 50,140,000 bp) harboured one protein coding gene (*PCDH7*) which coincides with a known CNV (Supplementary Table S4), one 5S ribosomal RNA (5S rRNA) and three non-coding RNA genes: SNORA70, Y_RNA and U6 (Table 1). *PCDH7* is one of the key genes involved in oncogenesis and/or differentiation of the cancer stem cells through a change in its histone methylation status[48]. Likewise, 5S rRNA genes are highly methylated in Arabidopsis thaliana and their expression is under epigenetic control[49,50].

During the process of fixation of adaptive variants, linked neutral markers are dragged along with the selected site; thus reducing the levels of genetic diversity in the region, while simultaneously new mutations accumulate in the region. The initial frequency of these mutations is low, so that a DNA sequence harbouring a positively selected variant will also harbour an excess of rare derived alleles. Bearing this in mind, we expect that many other sweeps are not detected by our genome scan owing to ascertainment schemes used to discover the BovineSNP50 BeadChip. Clearly, shedding light on additional selective sweeps in North African cattle would require the use of whole genome sequence data and the inclusion of all variants in genetic analyses.

The present study highlighted, for the first time, the presence of signatures of putative selection signatures in six local North African cattle populations. Information about the location of these regions can now be used as a starting point to identify causal genetic variants that control some environmental adaptation traits in local breeds which can be utilized in the genetic improvement of commonly used commercial breeds world-wide. Our results are unique in indicating that selection have shaped North African cattle genome through extensive regulation of gene expression whereby the individuals get adapted to short as well as long-term environmental changes. Understanding the functional consequences of such adaptive elements remains a challenge to overcome.

Methods

Data merging and SNP filtering

We combined Illumina BovineSNP50 BeadChip genotypes of 57 Brune de l'Atlas individuals (TUNIND) sampled from our previously published data[4,51] with data already available for 221 animals belonging to seven North African populations (BAL, BIS, CHE, CHF, GUE, TID and OUL) obtained from Flori *et al.,* 2018[16] and Gautier *et al.,* 2009[52].

We performed a relatedness test between individuals within each population using PLINK[53]. The software calculates a variable called PIHAT reflecting extended haplotypes shared between distantly related individuals. One individual from any pairs showing a PIHAT score ≥0.1 was removed from further analysis. After relatedness filtering, 207 North African individuals, including 39 TUNIND animals, were kept for population structure analyses. As detailed in Supplementary Table S10, we also included genotyping data belonging to 9 other populations, representatives of European taurines (EUT) (four breeds: ANG, HOL, JER and MON), African taurines (AFT) (three N'Dama populations: ND1, ND2 and NDA) and indicine (two populations: GIR and NEL) from Matukumalli *et al.*[54]. All genotypes were recovered from the web-interfaced genetic Diversity Exploration (WIDDE) database[55] and no relatedness analysis was performed for these breeds. We then conducted a series of quality control procedures. First, we excluded rare SNPs with low minor allele frequencies (MAF) < 0.05. Then, the whole genotype dataset was subjected to linkage disequilibrium (LD) pruning using the default parameters of PLINK (SNP window size :50, step 5 SNPs, r^2: 0.5). The combined dataset consisted of 494 individuals from 17 populations genotyped for 38,464 SNPs spread over all autosomal chromosomes.

Population structure and genetic relationship analyses

Population structure was inferred by PCA for African, European, indicine and North African populations using the adegenet R package[56]. Unsupervised hierarchical clustering was carried out for all populations using ADMIXTURE 1.23 software[57]. We ran ADMIXTURE with cross-validation for values of K from 2 through 17 (the number of populations) to identify the best value

of K clusters. DISTRUCT software[58] was then used to graphically display ancestry within each individual. The pairwise fixation index (F_{ST}) between populations was estimated using Genepop 4.6 software[59]. The patterns of population splits and mixtures were inferred using TreeMix[60]. First, we built a maximum likelihood tree of the 17 populations of the study with no migration events allowed and setting GIR as outgroup. Then, we built a phylogenetic tree of these populations and started adding migration events (modeled as edges) sequentially to the phylogenetic model. The migration edges were added until 99.92% of the variance in ancestry between populations was explained by the model. The residuals from the fit of the model to the data were visualized using the R script implemented in TreeMix.

Identification of selection signatures

To perform selection signature detection, we selected the individuals that are most representative of the ancestral North African cattle. This was done based on the results of model-based clustering results. We used the population differentiation based analysis implemented in BayeScan (F_{ST})[61] and three extended haplotype homozygosity (EHH)-based tests (*iHS*, *Rsb* and *XP-EHH*) to detect signatures of selection within North African cattle. Bayescan, *Rsb* and *XP-EHH* analyses were performed for each of the three pairwise comparisons: North African cattle Vs AFT, North African cattle Vs EUT and North African cattle Vs IND. Bayescan uses a reversible-jump Markov Chain Monte Carlo to separate locus-specific effects of selection from population-specific effects of demography. Outliers are those loci that require the locus-specific component to explain observed genetic diversity. For the Markov chain Monte Carlo (MCMC) algorithm we used 20 pilot runs of 5,000 iterations, a burn-in of 50,000 iterations, a thinning interval of 10 (5,000 iterations were used for the estimation of posterior odds) with a resulting total number of 100,000 iterations. To control the number of false positives, significant SNPs were defined by applying a *q*-value threshold of 0.05.

Haplotype extended patterns were then investigated using three metrics implemented in *rehh* package[62]: *iHS*[63], *Rsb* between pairs of populations[64] and *XP-EHH* within population[65]. In *iHS*

computation, the information on the ancestral and derived allele state is needed for each SNP because this statistic is based on the ratio of the EHH associated to each allele. In our analysis, the ancestral allele was inferred as the most common allele within 3 out-group species including yak, buffalo and sheep. *iHS* scores for each SNP were transformed into two-sided p-values : p*iHS*=−log10[1-2|Φ(*iHS*)-0.5|]. As a prerequisite to the *Rsb* and *XP-EHH* computation, haplotypes were reconstructed from the genotyped SNPs using fastPHASE 1.4[66]. The following options were used for each chromosome: -T10 -Ku60 -Kl10 -Ki10. Considering that *Rsb* and *XP-EHH* values are normally distributed, a Z-test was applied to identify significant SNPs under selection. Two-sided *p*-values were derived as p*Rsb*=−log10[1-2|Φ(*Rsb*)-0.5|] and p*XP-EHH*=−log10[1-2|Φ(*XP-EHH*)-0.5|] where Φ (x) represents the Gaussian cumulative distribution function. In EHH-based tests, the maximum allowed gap between two SNPs was set to 500 Kb. We used 1-Mb sliding windows that partially overlapped 10 kb with adjacent windows to perform selection signature detection. A window is classified as putatively under selection when it contains at least 3, 4 and 4 markers exceeding the significance threshold of −log10 (p-value) = 3 for *iHS, Rsb* and *XP-EHH* tests, respectively. Finally, we checked the overlap of the candidate genomic regions detected with at least two approaches with the previously identified bovine Quantitative Trait Loci (QTL) available in the cattle QTL database (http://www.animalgenome.org/cgi-bin/QTLdb/BT/index). The overlaps were checked using QTL coordinates according to the Bos taurus genome assembly ARS-UCD1.2.

Gene identification and functional enrichment analysis

Candidate genome region intervals detected by at least two methods (*iHS*, *Rsb*, *XP-EHH* or Bayescan) were interrogated for genes annotated to the *Bos taurus* genome assembly ARS-UCD1.2 using BioMart tool of Ensembl (https://www.ensembl.org/biomart/martview/c8fe3a69961a4088a55b7a249db7e2fa). Cattle structural variants which overlapped the genomic coordinates (in bp) of the candidate selective sweep regions were retrieved using the same database. We have only considered structural variants of less than 8 Mb which corresponds to the maximum size that can be identified, from whole

genome sequence data, by the pindel software (http://gmt.genome.wustl.edu/packages/pindel/user-manual.html). Functional annotation clustering was performed for the list the retrieved genes, using the Database for Annotation, Visualization and Integrated Discovery (DAVID) software version 6.8 (https://david.ncifcrf.gov/). DAVID was used to identify Gene Ontology (GO) terms with Benjamini-corrected p-value < 0.05.

References

1. Payne, W. J. A. & Hodges, J. Tropical cattle: origins, breeds and breeding policies. *Tropical cattle: origins, breeds and breeding policies*. (1997).
2. Decker, J. E. *et al.* Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genet* **10**, (2014).
3. Cymbron, T., Loftus, R. T., Malheiro, M. I. & Bradley, D. G. Mitochondrial sequence variation suggests an African influence in Portuguese cattle. *Proc Biol Sci* **266**, 597–603 (1999).
4. Ben Jemaa, S., Boussaha, M., Ben Mehdi, M., Lee, J. H. & Lee, S.-H. Genome-wide insights into population structure and genetic history of tunisian local cattle using the illumina bovinesnp50 beadchip. *BMC Genomics* **16**, (2015).
5. Ben Jemaa, S. *et al.* Genomic characterization of Algerian Guelmoise cattle and their genetic relationship with other North African populations inferred from SNP genotyping arrays. *Livestock Science* **217**, 19–25 (2018).
6. Boushaba, N. *et al.* Genetic diversity and relationships among six local cattle populations in semi-arid areas assessed by a bovine medium-density single nucleotide polymorphism data. *animal* **13**, 8–14 (2019).
7. Joshi, N. R., McLaughlin, E. A. & Phillips, R. W. *Types and Breeds of African Cattle*. (Food and Agriculture Organization of the United Nations, 1957).
8. Curson H.H & Thornton R.W. A contribution to the study of African native cattle. *Onderstepoort Journal of Veterinary Science and Animal Industry* **7**, 613–739 (1936).
9. Smith, J. M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genetics Research* **23**, 23–35 (1974).
10. Mei, C. *et al.* Genetic Architecture and Selection of Chinese Cattle Revealed by Whole Genome Resequencing. *Mol. Biol. Evol.* **35**, 688–699 (2018).
11. Mwacharo, J. M. *et al.* Genomic footprints of dryland stress adaptation in Egyptian fat-tail sheep and their divergence from East African and western Asia cohorts. *Scientific Reports* **7**, 1–10 (2017).

12. Ablondi, M., Viklund, Å., Lindgren, G., Eriksson, S. & Mikko, S. Signatures of selection in the genome of Swedish warmblood horses selected for sport performance. *BMC Genomics* **20**, 717 (2019).
13. Alshawi, A., Essa, A., Al-Bayatti, S. & Hanotte, O. Genome Analysis Reveals Genetic Admixture and Signature of Selection for Productivity and Environmental Traits in Iraqi Cattle. *Front Genet* **10**, 609 (2019).
14. Tijjani, A., Utsunomiya, Y. T., Ezekwe, A. G., Nashiru, O. & Hanotte, O. Genome Sequence Analysis Reveals Selection Signatures in Endangered Trypanotolerant West African Muturu Cattle. *Front. Genet.* **10**, (2019).
15. Rothammer, S., Seichter, D., Förster, M. & Medugorac, I. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics* **14**, 908 (2013).
16. Flori, L. *et al.* A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular Ecology* **28**, 1009–1029 (2019).
17. Blench, R. Ethnographic and linguistic evidence for the prehistory of African ruminant livestock, horses and ponies. *The archaeology of Africa. Food, metals and towns*.
18. da Fonseca, R. R. *et al.* Consequences of breed formation on patterns of genomic diversity and differentiation: the case of highly diverse peripheral Iberian cattle. *BMC Genomics* **20**, (2019).
19. Mastrangelo, S. *et al.* Genome-wide detection of signatures of selection in three Valdostana cattle populations. *Journal of Animal Breeding and Genetics* **n/a**,.
20. Beaumont, M. A. Adaptation and speciation: what can Fst tell us? *Trends in Ecology & Evolution* **20**, 435–440 (2005).
21. Excoffier, L., Hofer, T. & Foll, M. Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285–298 (2009).
22. Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M. & Servin, B. Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations. *Genetics* **193**, 929–941 (2013).
23. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).

24. Schrider, D. R., Mendes, F. K., Hahn, M. W. & Kern, A. D. Soft Shoulders Ahead: Spurious Signatures of Soft and Partial Selective Sweeps Result from Linked Hard Sweeps. *Genetics* **200**, 267–284 (2015).
25. Garud, N. R., Messer, P. W., Buzbas, E. O. & Petrov, D. A. Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. *PLOS Genetics* **11**, e1005004 (2015).
26. de Villemereuil, P., Frichot, É., Bazin, É., François, O. & Gaggiotti, O. E. Genome scan methods against more complex models: when and how much should we trust them? *Mol. Ecol.* **23**, 2006–2019 (2014).
27. The Bovine HapMap Consortium. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* **324**, 528–532 (2009).
28. Qanbari, S. *et al.* Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. *PLoS Genet* **10**, (2014).
29. Waszak, S. M. *et al.* Systematic Inference of Copy-Number Genotypes from Personal Genome Sequencing Data Reveals Extensive Olfactory Receptor Gene Content Diversity. *PLOS Computational Biology* **6**, e1000988 (2010).
30. Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **102**, 4501–4524 (2002).
31. Kim, J. *et al.* The genome landscape of indigenous African cattle. *Genome Biology* **18**, 34 (2017).
32. Gharbi, M., Rjeibi, M. R. & Darghouth, M. A. Epidémiologie de la theilériose tropicale bovine (infection par Theileria annulata) en Tunisie : une synthèse. *Revue d'élevage et de médecine vétérinaire des pays tropicaux* **67**, 241–247 (2014).
33. Rahali, T. *et al.* Séroprévalence et facteurs de risque des hémoparasitoses (theilériose, babésiose et anaplasmose) chez les bovins dans quatre grandes régions d'élevage du Maroc. *Revue d'élevage et de médecine vétérinaire des pays tropicaux* **67**, 235–240 (2014).
34. Geistlinger, L. *et al.* Widespread modulation of gene expression by copy number variation in skeletal muscle. *Scientific Reports* **8**, 1399 (2018).
35. Millar, D. S. *et al.* Growth hormone (GH1) gene variation and the growth hormone receptor (GHR) exon 3 deletion polymorphism in a West-African population. *Molecular and Cellular Endocrinology* **296**, 18–25 (2008).

36. Xiong, M.-G. *et al.* RNF152 positively regulates TLR/IL-1R signaling by enhancing MyD88 oligomerization. *EMBO reports* **21**, e48860 (2020).
37. Jeong, C., Nakagome, S. & Rienzo, A. D. Deep History of East Asian Populations Revealed Through Genetic Analysis of the Ainu. *Genetics* **202**, 261–272 (2016).
38. Pitt, D. *et al.* Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. *Evolutionary Applications* **12**, 105–122 (2019).
39. Kordonowy, L. *et al.* Physiological and biochemical changes associated with acute experimental dehydration in the desert adapted mouse, Peromyscus eremicus. *Physiol Rep* **5**, (2017).
40. Ahlawat, S. *et al.* Skin transcriptome profiling of Changthangi goats highlights the relevance of genes involved in Pashmina production. *Scientific Reports* **10**, 6050 (2020).
41. Yang, C.-C. & Cotsarelis, G. Review of hair follicle dermal cells. *J Dermatol Sci* **57**, 2 (2010).
42. Tamate, S. *et al.* Inferring evolutionary responses of Anolis carolinensis introduced into the Ogasawara archipelago using whole genome sequence data. *Scientific Reports* **7**, 18008 (2017).
43. Mahootchi, E. *et al.* GADL1 is a multifunctional decarboxylase with tissue-specific roles in β-alanine and carnosine production. *Science Advances* **6**, eabb3713 (2020).
44. Niu, G. *et al.* Comparative and evolutionary analyses of the divergence of plant oligosaccharyltransferase STT3 isoforms. *FEBS Open Bio* **10**, 468–483 (2020).
45. Sierra, B. *et al.* OSBPL10, RXRA and lipid metabolism confer African-ancestry protection against dengue haemorrhagic fever in admixed Cubans. *PLoS Pathog* **13**, (2017).
46. Wagnon, J. L. *et al.* CELF4 Regulates Translation and Local Abundance of a Vast Set of mRNAs, Including Genes Associated with Regulation of Synaptic Function. *PLOS Genetics* **8**, e1003067 (2012).
47. Yang, Y. *et al.* Complex Seizure Disorder Caused by Brunol4 Deficiency in Mice. *PLOS Genetics* **3**, e124 (2007).
48. Zhang, Q. *et al.* Mdig promotes oncogenic gene expression through antagonizing repressive histone methylation markers. *Theranostics* **10**, 602–614 (2020).
49. Simon, L. *et al.* Genetic and epigenetic variation in 5S ribosomal RNA genes reveals genome dynamics in Arabidopsis thaliana. *Nucleic Acids Res* **46**, 3019–3033 (2018).
50. Pontvianne, F. *et al.* Histone methyltransferases regulating rRNA gene dose and dosage control in Arabidopsis. *Genes Dev* **26**, 945–957 (2012).

51. Ben Jemaa, S. *et al.* Linkage disequilibrium and past effective population size in native Tunisian cattle. *Genet Mol Biol* **42**, 52–61 (2019).
52. Gautier, M. *et al.* A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* **10**, 550 (2009).
53. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* **81**, 559–575 (2007).
54. Matukumalli, L. K. *et al.* Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS One* **4**, (2009).
55. Sempéré, G. *et al.* WIDDE: a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC Genomics* **16**, (2015).
56. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
57. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
58. Rosenberg, N. A. distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137–138 (2004).
59. Rousset, F. genepop'007: a complete re- implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103–106 (2008).
60. Pickrell, J. K. & Pritchard, J. K. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet* **8**, (2012).
61. Foll, M. & Gaggiotti, O. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **180**, 977–993 (2008).
62. Gautier, M. & Vitalis, R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177 (2012).
63. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
64. Tang, K., Thornton, K. R. & Stoneking, M. A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS Biol* **5**, (2007).

65. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
66. Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* **78**, 629–644 (2006).

Author contributions

S.B.J and M.B conceived the study. S.B.J, S.M and M.B performed the analysis. S.H.L and J.H.L contributed to data acquisition. S.B.J wrote the manuscript. S.B.J, S.M and M.B revised the manuscript.

Acknowledgements

This study was supported by grants from International Foundation for Science (IFS grant B/5478), and by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT & Future Planning (2013K2A4A1044519). The authors would like to thank Dominique Rocha and Emmanuelle Rebours (GABI unit) for their help in data acquisition and genotyping.

Competing Interests: The author(s) declare no competing interests.

Figure legends

Figure 1. Principle component analysis results of allele frequencies obtained from 38,464 SNPs genotyped in 494 cattle individuals from 17 populations. Each point represents the eigenvalues of principal components 1 and 2. Populations are represented by coloured inertia ellipses.

Figure 2. Unsupervised hierarchical clustering of the 494 individuals from the 17 populations of the study. Results for K (number of clusters) = 2, 3, 5, 7, 10, 12 (k-value with the lowest cross-validation error) and 17 are shown. Individuals are grouped by population. Each individual is represented by a vertical bar. The proportion of the bar in each of K colours corresponds to the average posterior likelihood that the individual is assigned to the cluster indicated by that colour. Populations are separated by black lines.

Figure 3. Maximum likelihood tree constructed with TreeMix when 14 migration events (modeled as arrows) were allowed. Migration arrows are coloured according to their weight.

Figure 4. Manhattan plots showing the results of *Rsb* test for the autosomes in North African cattle. **(a)** Rsb test AFT Vs North African cattle. **(b)** Rsb test EUT Vs North African cattle. **(c)** Rsb test IND Vs North African cattle. Horizontal dashed lines mark the significance threshold applied to detect the outlier SNPs (–log10 (p-value) = 3.

Figure 5. Manhattan plots showing the results of *XP-EHH* and *iHS* tests for the autosomes in North African cattle. **(a)** *XP-EHH* test AFT Vs North African cattle. **(b)** *XP-EHH* test EUT Vs North African cattle. **(c)** *XP-EHH* test IND Vs North African cattle. **(d)** iHS test for North African cattle. Horizontal dashed lines mark the significance threshold applied to detect the outlier SNPs (–log10 (p-value) = 3.