



OPEN

BoLA-DRB3 gene haplotypes show divergence in native Sudanese cattle from taurine and indicine breeds

Bashir Salim^{1✉}, Shin-nosuke Takeshima², Ryo Nakao³, Mohamed A. M. Moustafa³, Mohamed-Khair A. Ahmed⁴, Sumaya Kambal⁵, Joram M. Mwacharo⁶, Abeer M. Alkhaibari⁷ & Guillermo Giovambattista⁸

Autochthonous Sudanese cattle breeds, namely Baggara for beef and Butana and Kenana for dairy, are characterized by their adaptive characteristics and high performance in hot and dry agro-ecosystems. They are thus used largely by nomadic and semi-nomadic pastoralists. We analyzed the diversity and genetic structure of the BoLA-DRB3 gene, a genetic locus linked to the immune response, for the indigenous cattle of Sudan and in the context of the global cattle repository. Blood samples (n = 225) were taken from three indigenous breeds (Baggara; n = 113, Butana; n = 60 and Kenana; n = 52) distributed across six regions of Sudan. Nucleotide sequences were genotyped using the sequence-based typing method. We describe 53 alleles, including seven novel alleles. Principal component analysis (PCA) of the protein pockets implicated in the antigen-binding function of the MHC complex revealed that pockets 4 and 9 (respectively) differentiate Kenana-Baggara and Kenana-Butana breeds from other breeds. Venn analysis of Sudanese, Southeast Asian, European and American cattle breeds with 115 alleles showed 14 were unique to Sudanese breeds. Gene frequency distributions of Baggara cattle showed an even distribution suggesting balancing selection, while the selection index (ω) revealed the presence of diversifying selection in several amino acid sites along the *BoLA-DRB3* exon 2 of these native breeds. The results of several PCA were in agreement with clustering patterns observed on the neighbor joining (NJ) trees. These results provide insight into their high survival rate for different tropical diseases and their reproductive capacity in Sudan's harsh environment.

There is a consensus among population geneticists that the Sudanese cattle populations belong to the humped Zebu cattle breed and are classified into two principal varieties: northern Sudan and Nilotic^{1,2}. The Kenana and Butana breeds are the best-known milk-producing northern Sudan Zebu breeds³⁻⁵ with milk yield of over 1500 kg per lactation⁶⁻⁸.

The Kenana breed, predominantly found in the Blue Nile state, is distinguished by a light blue-gray coat color with darker hooves and head. The Butana breed of the Batahin and Shukria tribes inhabits the desert area between the Blue Nile and the River Atbara, and has a red-coat¹. A third breed, the Baggara, is raised by Baggara Bedouin pastoralists. It is the major fattening Zebu cattle breed of northern Sudan, found mostly in west Sudan (Darfur and Kordofan regions), Niger, Chad, Cameroon and Nigeria. They have short horns and a large hump, with a red or dark red coat in Daeinawi Aka Messairi/Rezaigi population or white markings or black markings in the Nyalawi population⁹.

¹Department of Parasitology, Faculty of Veterinary Medicine, University of Khartoum, Khartoum-North, Sudan. ²Department of Food and Nutrition, Faculty of Human Life, Jumonji University, 2-1-28 Sugawara, Niiza, Saitama, Japan. ³Laboratory of Parasitology, Department of Disease Control, Graduate School of Veterinary Medicine, Hokkaido University, Sapporo, Japan. ⁴Department of Genetics and Animal Breeding, Faculty of Animal Production, University of Khartoum, Khartoum, Sudan. ⁵National University, Biomedical Research Institute, Khartoum, Sudan. ⁶International Centre for Agricultural Research in the Dry Areas (ICARDA), Beirut, Lebanon. ⁷Department of Biology, Faculty of Science, University of Tabuk, Tabuk 71491, Kingdom of Saudi Arabia. ⁸Facultad de Ciencias Veterinarias UNLP, IGEVET - Instituto de Genética Veterinaria (UNLP-CONICET LA PLATA), La Plata, Argentina. ✉email: bashirsalim@gmail.com

The immune system in vertebrates evolved to defend against invasive pathogens¹⁰ and thus it is not surprising that genetic factors are implicated in disease susceptibility in cattle. The major histocompatibility complex (MHC) is a major component of the adaptive immune system, with MHC genes encoding the cell-surface glycoproteins that bind small peptide fragments derived from host- and pathogen-expressed proteins via proteolysis¹¹. Animal breeders are becoming more interested in the MHC due to its association with genetic resistance and susceptibility to a wide variety of diseases¹². Genetic characterization of MHC polymorphism can help reduce the occurrence and severity of infectious diseases in domestic animal and cattle breeding programs¹³. The association of MHC with diseases in ruminants is well documented^{14–19}. The MHC genes are assigned to *Bos taurus* autosome chromosome 23 (BTA 23)^{14,15}, and is known as the Bovine Leukocyte Antigen (BoLA). Recently Kim et al.¹⁶, investigated five African breeds for the identification of common and unique African genome-specific selection signatures and compared them with commercial breeds. They identified six BoLA haplotype blocks, and that the major African cattle haplotypes correspond to minor haplotypes in commercial cattle. The BoLA molecules' extensive structural polymorphism is responsible for the large differences in cattle's immune response to infectious agents. For example, *BoLA-DRB3* polymorphisms had been associated with tick infestation resistance¹⁶. MHC research may also assist in the formation and design of synthetic peptide-based vaccines containing one or more pathogen T-cell epitope.

Polymerase chain reaction-sequence based typing (PCR-SBT) to assess the genetic diversity of the *BoLA-DRB3* gene has been done with only a few breeds and cross-breeds from Europe, Asia and the Americas^{17–30}. Target next generation sequencing (Target-NGS)³¹, the most powerful tools used to identify diversity of *BoLA-DRB3* alleles in cattle breeds, has also not been extensively applied. This is despite the central role of the *BoLA-DRB3* alleles in the immune response of cattle. Until now, private African *BoLA-DRB3* alleles have been reported by authors using indirect techniques, such as polymerase chain reaction followed by restriction fragment length polymorphism (PCR-RFLP), followed by cloning and sequencing^{32–34}. These studies focused mainly on screening and analysis of only a few animals from a small selection of African breeds (e.g. Sanga, Kenana, Butana).

Previous work showed the presence of a high number of private alleles in native breeds. Consequently, there are still a number of breeds that remain uncharacterized, and this number only increases when local native bovine breeds are considered^{19,21–23}.

Here we examine patterns of genetic variation of *BoLA-DRB3* alleles in Baggara, Butana and Kenana native cattle breeds of Sudan and compare these with commercial breeds to both identify any unique alleles in Sudanese native cattle breeds and to provide information on these alleles to enable further studies of disease susceptibility and resistance, particularly for designing improved breeding schemes.

Results

Distribution of *BoLA-DRB3* alleles in selected native Sudanese cattle breeds. PCR-SBT genotyping allowed us to identify 53 *BoLA-DRB3* alleles (46 previously reported variants and seven new alleles; Table 1) from the native breeds selected in this study. The number of alleles (n_a) was 46 in Baggara cattle (40 previously reported and six new), 33 in Kenana cattle (28 previously reported and five new), and 33 in Butana cattle (28 previously reported and five new) (Tables 1 and 2). The new *BoLA-DRB3* variants were confirmed by the presence of at least three carrier animals and in two breeds, and were submitted to the DNA Data Bank of Japan (<http://www.ddbj.nig.ac.jp>) under accession numbers LC569724–LC569739. Nucleotide and predicted amino acid sequences of the seven new allele variants are shown in Fig. 1 and compared with the most similar *BoLA-DRB3* reported so far. All seven new *BoLA-DRB3* allele variants shared about 89.7–92.6% and 80.52–85.71% nucleotide and amino acid similarity with the *BoLA-DRB3* cDNA clone NR1, respectively (Aida, 1995).

A Venn diagram was constructed using data obtained in this study and from previous reports^{18,19,21,27,29}. Data were grouped in terms of the breed's geographical origin as follows: native Sudanese; Southeast Asian; Zebu; European; and American Creole cattle breeds (Fig. 2). This analysis revealed that out of the 115 alleles identified in the five cattle groups, fourteen were unique to native Sudanese breeds (Fig. 2), four of which exhibited gene frequencies that were higher than 0.5%, representing about 26% of the 53 alleles detected in the native Sudanese cattle. In addition, two other variants were only present in native Sudanese and American Creole breeds, while six other alleles were only found in Sudanese cattle populations and American Creole or Southeast Asian native or Zebu breeds, or a combination of these groups. In addition, the *BoLA-DRB3* NJ tree, including all the previously reported alleles and the seven new variants, showed that the variants detected in Sudanese cattle populations were interspersed among the various clusters (Fig. 3). A similar result was observed when the *BoLA-DRB3* tree was inferred using amino-acid residues located in the antigen-binding site (ABS) (Fig. S1).

As shown in Fig. S2, the native Sudanese cattle breeds have an even gene frequency distribution, with a high number of alleles with low frequency. Low allele frequency was particularly noticeable in the Baggara breed. Only two, five and seven alleles appeared with frequencies of > 5% in the Baggara, Kenana and Butana breeds, respectively. These common alleles accounted for a low proportion of the cumulative gene frequencies (12.83, 44.23 and 50.83% in the Baggara, Kenana and Butana breeds, respectively); four of which (*BoLA-DRB3**003:02:01, *021:01 *022:01 and *024:01) were common in at least two out of the three Sudanese breeds (Table 1).

Nucleotide and amino acid diversity in the *BoLA-DRB3* alleles found in native Sudanese cattle breeds. Genetic diversity at the DNA and amino acid levels was evaluated using four methods that compare the average amino acid and nucleotide substitutions for every pair of alleles within the breeds. The nucleotide diversity (π) exceeded 0.074 and the mean number of pairwise differences values exceeded 17.99 within Sudanese native breeds (Table 3). Comparison with results previously reported for other cattle breeds showed that these nucleotide diversity values all fall within the range previously reported ($\pi_{\text{range}} = 0.068–0.090$; $\text{NPD}_{\text{range}} = 16.31–20.96$) when using PCR-SBT genotyping methods^{18,19,21,29,30}. Regarding amino acid diversity,

DRB3	Baggara	Kenana	Butana
	(N = 113)	(N = 52)	(N = 60)
<i>BoLA-DRB3*003:01</i>	0.00	0.96	0.00
<i>BoLA-DRB3*003:02:01</i>	4.87	7.69	5.83
<i>BoLA-DRB3*004:01</i>	2.65	0.00	0.00
<i>BoLA-DRB3*005:02</i>	4.87	2.88	2.50
<i>BoLA-DRB3*007:01</i>	2.65	0.96	0.00
<i>BoLA-DRB3*008:01</i>	1.33	0.00	0.83
<i>BoLA-DRB3*009:01</i>	0.88	0.00	0.00
<i>BoLA-DRB3*009:02</i>	2.65	0.96	4.17
<i>BoLA-DRB3*010:01</i>	0.44	0.96	4.17
<i>BoLA-DRB3*010:02</i>	1.33	1.92	1.67
<i>BoLA-DRB3*011:01</i>	4.42	0.00	0.00
<i>BoLA-DRB3*011:02</i>	0.44	0.00	0.00
<i>BoLA-DRB3*012:01</i>	2.65	3.85	0.00
<i>BoLA-DRB3*013:01</i>	3.54	0.00	1.67
<i>BoLA-DRB3*014:01:01</i>	4.42	2.88	9.17
<i>BoLA-DRB3*015:01</i>	4.42	1.92	0.83
<i>BoLA-DRB3*016:01</i>	5.75	0.00	0.83
<i>BoLA-DRB3*018:01</i>	2.21	4.81	4.17
<i>BoLA-DRB3*019:01</i>	0.00	0.00	0.83
<i>BoLA-DRB3*020:01:01</i>	1.32	4.81	0.83
<i>BoLA-DRB3*020:02</i>	0.88	0.00	0.00
<i>BoLA-DRB3*020:03</i>	0.44	0.00	0.83
<i>BoLA-DRB3*021:01</i>	3.10	7.69	10.83
<i>BoLA-DRB3*022:01</i>	7.08	6.73	9.17
<i>BoLA-DRB3*022:10</i>	0.00	0.00	0.83
<i>BoLA-DRB3*022:12</i>	0.44	0.00	0.00
<i>BoLA-DRB3*023:01</i>	1.77	2.88	2.50
<i>BoLA-DRB3*024:01</i>	4.87	14.42	5.00
<i>BoLA-DRB3*024:06</i>	0.00	1.92	0.83
<i>BoLA-DRB3*025:01:01</i>	0.44	0.96	0.00
<i>BoLA-DRB3*026:01</i>	2.65	0.96	3.33
<i>BoLA-DRB3*027:03</i>	0.44	0.96	0.00
<i>BoLA-DRB3*027:04</i>	1.33	1.92	1.67
<i>BoLA-DRB3*027:05</i>	0.00	0.96	0.83
<i>BoLA-DRB3*028:01</i>	3.98	7.69	3.33
<i>BoLA-DRB3*028:02</i>	3.10	0.96	4.17
<i>BoLA-DRB3*029:02</i>	0.44	0.00	0.00
<i>BoLA-DRB3*030:01</i>	0.44	0.00	0.00
<i>BoLA-DRB3*032:03</i>	0.00	0.96	0.00
<i>BoLA-DRB3*033:01</i>	0.44	1.92	0.83
<i>BoLA-DRB3*036:01</i>	1.33	0.00	0.00
<i>BoLA-DRB3*039:01</i>	0.44	0.96	1.67
<i>BoLA-DRB3*044:01</i>	1.77	0.00	1.67
<i>BoLA-DRB3*048:02</i>	4.42	1.92	0.00
<i>BoLA-DRB3*100:01</i>	0.44	0.00	0.00
<i>BoLA-DRB3*107:01</i>	0.44	0.00	0.00
<u><i>BoLA-DRB3*004:02sp2</i></u>	0.44	0.00	5.00
<u><i>BoLA-DRB3*011:02Sp</i></u>	3.10	2.88	5.83
<u><i>BoLA-DRB3*018:01Sp</i></u>	0.44	1.92	0.00
<u><i>BoLA-DRB3021:01sp</i></u>	0.88	0.00	1.67
<u><i>BoLA-DRB3*024:18Sp</i></u>	0.00	3.85	0.83
<u><i>BoLA-DRB3*027:05Sp</i></u>	2.21	1.92	1.66
<u><i>BoLA-DRB3*032:01Sp</i></u>	1.33	0.96	0.00

Table 1. BoLA-DRB3 allele frequencies (in percentage) in native Sudanese cattle breeds. N. number of animals analyzed; Frequent alleles in each breed are indicated in bold (> 5%); Novel alleles identified in this study are indicated in bold and underlined.

Breed	N	n _a	New alleles	h _o	h _e	HWE	
						F _{IS} -p value	Slatkin's p value
Baggara	113	46	6	0.938	0.969	0.032–0.277	0.016
Butana	52	33	5	0.961	0.951	0.005–0.100	0.255
Kenana	60	33	5	0.950	0.954	-0.011–0.628	0.138

Table 2. Number of alleles (n_a) and new alleles, observed (h_o) and expected (h_e) heterozygosity, Hardy Weinberg equilibrium (HWE) measured through F_{IS} and Slatkin's exact test in the cattle breeds studied. N = sample size.

the average d_N and d_S substitutions in Sudanese cattle breeds were calculated across the entire *BoLA-DRB3* exon 2 and ABS. As expected, the d_N/d_S ratio was higher when only the ABS was analyzed (Table 3). These values obtained in Sudanese cattle were similar to those estimated for other cattle breeds (d_N/d_S total = 3.50–3.85; d_N/d_S ABS = 4.80 – 4.93).

Gene diversity, Hardy–Weinberg Equilibrium (HWE), and neutrality testing of *BoLA-DRB3* variants found in Sudanese cattle breeds.

Genetic diversity within the three Sudanese breeds was estimated using the n_a and gene diversity (h_o and h_e). We also performed HWE and Slatkin's neutrality tests on *BoLA-DRB3* to evaluate the possible effect of selection, inbreeding, and population structure on allelic diversity at this locus. The high n_a values and even gene frequencies observed in the Butana, Kenana and Baggara breeds resulted in h_e and h_o values higher than 0.93 (Table 2). As expected, these indices highlighted extremely high diversity values for Sudanese cattle populations, which is similar to the results reported for other bovine breeds which have been evaluated by PCR-SBT, and characteristic of *MHC class II DR* genes^{18,19,21,27,29,35}. Regarding the HWE test, the three Sudanese native populations were in equilibrium (Table 2), similar to observations in half of the bovine breeds studied so far. It is widely accepted that the genetic diversity of *MHC class II* genes can be maintained by balancing selection. Thus, we performed a Slatkin's exact neutrality test (Table 2) to evaluate this phenomenon in the Sudanese cattle populations. The *BoLA-DRB3* gene frequency profile in Baggara cattle showed an even distribution ($p=0.016$), consistent with the theoretical proportion expected under balancing selection pressures. A similarly even *BoLA-DRB3* gene frequency was observed in other cattle breeds, including Japanese Black, Yacumeño Creole, Bolivian Gir, Pyer Sein and Shwe Ni. Conversely, we did not detect balancing selection in the Butana and Kenana cattle ($p=0.225$ and $p=0.138$) despite these breeds having a large number of alleles with similar frequency. Comparable results were obtained for the majority of the cattle breeds analyzed to date (Table 2). In addition, we estimated the selection index (ω) in each amino acid site to evaluate the presence of diversifying selection ($\omega > 1$) along *BoLA-DRB3* exon 2. These analyses showed high ω values in more than 30 sites in each breed, mainly located in the ABS (Fig. 4).

BoLA-DRB3 genetic structure and levels of population differentiation in Sudanese cattle.

The level of genetic differentiation among the three Sudanese breeds was studied through the F_{ST} index. The average F_{ST} was statistically significant although this value accounts for less than one percent of the total genetic variance ($F_{ST}=0.0076$ (ranging between 0.007 and 0.009); $p < 0.001$) (Table S2). This low but significant value can be explained by high within-population diversity and differences in rare allele profiles among them³⁶. The average F_{ST} value observed in Sudanese cattle is higher than those estimated in Myanmar native breeds ($F_{ST}=0.003$), and slightly lower than those reported for Holstein populations from different countries ($F_{ST}=0.009$)^{18,37} (Fig. 5 and Table S2). When breeds were grouped in terms of the breed's geographical origin, as was done in the Venn diagram, the genetic variance among breed groups and among populations within groups accounted for 1.18% and 3.71% of the total genetic variance. Table S2 summarizes the genetic distance, measured by F_{ST} , between native Sudanese breeds and other taurine and zebu breeds for *BoLA-DRB3*, showing that native Sudanese cattle diverge from other breeds with F_{ST} values between 0.014 and 0.082.

When the five sampling sites of native Sudanese breeds were compared (two sampling locations of Kenana cattle were very close and assumed as one), the average F_{ST} value was 0.0074 ($p=0.164$), while the pairwise F_{ST} ranged from 0.0002 ($p=0.450$) between both Baggara populations and 0.0118 ($p < 0.0001$) between Baggara Daiwani and Butana Qadarif. Significant differences were observed in nine out of the ten native population comparisons ($p < 0.05$; Table S3). Similar genetic distance values were observed among Holstein populations from different countries and between native breeds of Myanmar^{18,37}.

Genetic differentiation of *BoLA-DRB3* alleles in native Sudanese cattle breeds: comparison with Zebu and Taurine breeds.

First, *BoLA-DRB3* allele frequencies from Sudanese cattle populations and for each breed included in the dataset were used to generate Nei's D_A and D_S genetic distance matrices. Then, dendrograms were constructed from these distance matrices using NJ algorithm. All trees revealed congruent topologies, which were consistent with the historical and geographical origin of the breeds analyzed. As expected, these trees revealed two main clusters, which included the Taurine and Zebuine breeds (Fig. 6a). It is noteworthy that Sudanese breeds were located in a sub-cluster within the indicine cluster, with the two dairy breeds located in the east of the country, Butana and Kenana being more related to each other than the Baggara

A	1	10	11	20	21	30	31	40	41	40	51	60	61	70	71	80
BoLA-DRB3*016:01	TCTCTCTGCA	GCACATTTCC	TGGAGTATAC	CAAGAAAGAG	TGTCATTCT	TCAACGGGAC	CGAGCGGGTG	CGGTTCTCTGG								
BoLA-DRB3*004:02Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*004:02	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*011:02Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*011:02	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*018:01Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*018:01	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*021:01sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*021:01	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*024:18Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*024:18	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*027:05Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*027:05	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*032:01Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*032:01	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

	81	90	91	100	101	110	111	120	121	120	131	140	141	150	151	160
BoLA-DRB3*016:01	ACAGATACTT	CCATAATGGA	GAAGATTTCG	TGCGCTTCGA	TAGCGACTGG	GGCGAGTACC	GGGCGGTGAC	CGAGCTAGGG								
BoLA-DRB3*004:02Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*004:02	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*011:02Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*011:02	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*018:01Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*018:01	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*021:01sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*021:01	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*024:18Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*024:18	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*027:05Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*027:05	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*032:01Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*032:01	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

	161	170	171	180	181	190	191	200	201	210	211	220	221	230	231	240
BoLA-DRB3*016:01	CGGCCGGACG	CCAAGTACTG	GAACAGCCAG	AAGGACTTCC	TGGAGGAGAA	GCGGGCCGCG	GTGGACACGT	ACTGCAGACA								
BoLA-DRB3*004:02Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*004:02	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*011:02Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*011:02	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*018:01Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*018:01	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*021:01sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*021:01	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*024:18Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*024:18	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*027:05Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*027:05	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*032:01Sp	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*032:01	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

	241	250	251	260	261	270	280	281
BoLA-DRB3*016:01	CAACTACGGG	GTCGGTGAGA	GTTTCACTGT	GCAGCGGCGA	G			
BoLA-DRB3*004:02Sp	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*004:02	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*011:02Sp	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*011:02	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*018:01Sp	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*018:01	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*021:01sp	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*021:01	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*024:18Sp	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*024:18	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*027:05Sp	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*027:05	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*032:01Sp	-----	-----	-----	-----	-----	-----	-----	-----
BoLA-DRB3*032:01	-----	-----	-----	-----	-----	-----	-----	-----

Figure 1. Alignment of the nucleotide (A) and the predicted amino acid (B) sequences of the $\beta 1$ domain encoded by seven new *BoLA-DRB3* alleles (accession numbers, LC569725 for *BoLA-DRB3*004:02Sp*, LC569726 for *BoLA-DRB3*011:02Sp*, LC569729 for *BoLA-DRB3*018:01Sp*, LC569731 for *BoLA-DRB3*021:01sp*, LC569733 for *BoLA-DRB3*024:18Sp*, LC569735 for *BoLA-DRB3*027:05sp*, and LC569739 for *BoLA-DRB3*032:01sp*) derived from 225 Sudan native cattle (113 animals of the Baggara native, 60 Butana, and 52 Kenana Sudan native cattle breeds). New alleles are indicated in bold. Numbering refers to amino acid positions in the mature protein. Nucleotide and amino acid residues identical to those encoded by the *BoLA-DRB3* cDNA clone NR-1 are indicated by dots (Aida et al., 1995). Missing data are indicated by dashes. Closer *BoLA-DRB3* alleles with new variants are also included in the figure. Id. = Nucleotide or amino acid identity in %.

B	6	
BoLA-DRB3*016:01	HFLEYTKKECHFFNGTERVRFRLDRYFHNGEFVRFDSDWGEYRAVTELGRPDAKYWNSSQKDFLEEKRAAVDTYCRHNYGVGESFTVQRR	
BoLA-DRB3*004:02Sp	---.STS.....Y.....Y.....F.....V..RE..Y.....GV.....-----	84.42
BoLA-DRB3*004:02	---.STS.....Y.....Y.....FQ.....I..RE..Y.....G.....-----	
BoLA-DRB3*011:02Sp	---.Y.G.....L..H.Y.....Y.....F.....E.....R..E..V.....V.....-----	84.42
BoLA-DRB3*011:02	---.Y.G.....L..H.Y.....Y.....F.....S.E.....R..E..V.....V.....QRR	
BoLA-DRB3*018:01Sp	---.ATS.....H.....Y.....Y.....F.....A.EQ.....HT..RE..Y.....GV.....-----	80.52
BoLA-DRB3*018:01	---.ATS.....H.....Y.....Y.....F.....A.EQ.....T..RE..Y.....GV.....QRR	
BoLA-DRB3*021:01sp	---.STG.....Y.....Y.....Y.....Q..EQ.....R..E.....V.....-----	85.71
BoLA-DRB3*021:01	---.STG.....Y.....Y.....Y.....Q..EQ.....R..E.....V.....QRR	
BoLA-DRB3*024:18Sp	---.C.R.....Y.....Y.....F.....S.E.....EI..R..E..RV.....-----	83.12
BoLA-DRB3*024:18	---.C.R.....Y.....Y.....F.....A.E.....EI..R..E..RV.....QRR	
BoLA-DRB3*027:05Sp	---.Y.R.....CYT..T.....F.....E.....R..E..RV.....V.....-----	83.12
BoLA-DRB3*027:05	---.Y.R.....CYT..T.....F.....E.....R..E..RV.....V.....---	
BoLA-DRB3*032:01Sp	---.C.R.....Y.....Y.....F.....E.....EI..R..Y.....V.....-----	85.71
BoLA-DRB3*032:01	---.C.R.....Y.....Y.....F.....E.....EI..RE..Y.....QRR	

Figure 1. (continued)

breed in the west. These results reveal that Sudanese cattle breeds have a particular diversity in the *BoLA-DRB3* gene, as a consequence of its gene frequency profile and the presence of a high number of private alleles.

The results of the PCA showed that the first three components accounted for 47.30% of the data variability. The first principal component (PC) accounted for 24.31% of the total variance and, as shown in a previous study⁶⁴, clearly exhibited a differentiation pattern between the Zebu (negative values) and Taurine (positive values) breeds, while native breeds from Southeast Asia and Sudan were located in an intermediate position near the axis origin of the plot (Fig. 6b). This PC was primarily determined by differences in the frequency of the same alleles, such as *BoLA-DRB3*022:01*, **028:01*, **036:01*, **031:01*, **030:01*, and **057:02* with the higher negative axis 1 values, whereas the alleles *BoLA-DRB3*001:01*, **002:01*, **007:01*, **008:01*, **010:01*, **011:01*, **012:01*, **015:01* **016:01*, **018:01* had the higher positive values for this axis. The second PC explained 11.98% of the total variation and showed a gradient among Taurine breeds, with Chilean Hereford (positive values) and Japanese Jersey (negative values) located at opposite ends. Furthermore, this component discriminated between native Sudanese and native Southeast Asian cattle breeds. Finally, the third PC accounted for 11.01% of the variance and allowed for the differentiation of Chilean Hereford, and Japanese Jersey and Japanese Holstein cattle from other Taurine breeds. In summary, the native Sudanese cattle breeds were located within a narrow cloud in an intermediate position between the Zebu and Taurine breeds and close to other Southeast Asian breeds, in agreement with the composite origin of these native breeds. This is also supported by the presence of African and Zebu unique *BoLA-DRB3* alleles within these populations. These PCA results agree with the overall clustering observed after NJ tree construction.

The BoLA class II molecule binds peptides derived from antigens via five antigen binding pockets named pocket 1, pocket 4, pocket 6, pocket 7 and pocket 9²⁴. To assess whether observed differences in allelic frequency are reflected within amino acid motifs in each pocket, we analyzed frequency of the protein pockets implicated in the antigen-binding function of the MHC complex by PCA. As shown in Fig. S3a-e, the three native breeds of Sudan are located in a closed cloud in the five PCAs made based on the frequency of the pockets, although varying their relative position with other breeds and breed groups, and in some cases the spatial distribution did not exhibit a clear relationship with the geographical or historical origin of the breeds. However, pockets 4 and 9 are the ones that best differentiate these native breeds from the rest. Regarding pocket 4, Baggara and Kenana breeds of Sudan are located in a narrow cloud located at the end of axis 2, and their position is mainly explained by the GFDEREY, RFDERFV and GLDRKEV motifs. The position of the Butana and Kenana Sudanese breeds in pocket 9 was the result of positive PC1 and PC2 values for the presence of amino acid motifs EYD and EFA.

Finally, PCA was performed at the Sudanese population level to evaluate the degree of genetic structure among the sampling sites (Baggara Daiwani, Baggara Nyakawi, Kenana, Butana Bu Atbara and Butana Bu Qadarif). This analysis showed that the first three components accounted for 90.95% of the data variability. The first PC accounted for 30.65% of the total variance and clearly exhibited a differentiation pattern between the Baggara population (negative values) and the Butana Bu Qadarif (positive values) population, while Kenana, Butana Bu Atbara were located in intermediate positions (Fig. 7). These results agree with the geographical distribution of the studied population. The second and third PCs explained 30.66% and 25.24% of the total variation and allowed for the differentiation of the Butana Qadarif and Kenana populations, respectively.

Discussion

Since the first pioneering studies based on serotype analysis, a number of striking differences between the BoLA profiles of African and European cattle have been reported due to difference in the antigen's frequency of occurrence and the presence of unique antigens in African cattle³⁸. Over the next decades, several private alleles were identified in taurine, zebu and taurindicus native African breeds, like N'Dama, Boran, and Sanga^{32,34}; <https://www.ebi.ac.uk/ipd/mhc/group/BoLA/>). However, in the present study, we carried out the first genetic characterization of the *BoLA-DRB3* gene at population level in native Sudanese breeds using PCR-SBT. This analysis allowed us to detect 53 alleles, including seven new variants. The high number of private alleles agrees with data obtained by¹⁶, who analyzed the BoLA region in depth using a genome-wide sequencing approach, identifying six major African BoLA haplotype blocks.

Wild cattle or 'aurochs' (*Bos primigenius*), the ancestor of domestic cattle, inhabited a large geographical area throughout Eurasia and North Africa. According to the trans-species theory of MHC alleles³⁹, it is expected that

the extremely high genetic variability present in the *BoLA-DRB3* gene (365 alleles have been reported in the IPD-MHC (<https://www.ebi.ac.uk/ipd/mhc/group/BoLA>;³³ database, access date 16/04/21) was present in the wide geographical distribution of the aurochs. On the basis of archeological and genetic studies, it has been proposed that modern bovines were domesticated in two geographical sites, one located in the West Asia (Near east), and the other in Indian subcontinent (India and Pakistan)^{40–46}. Each of these domestication centers would have retained only a fraction of the total diversity as a result of bottleneck and genetic drift effects⁴⁷. This is clearly seen in the distribution of mitochondrial haplogroups among cattle breeds^{5,40–44}. In Africa, taurine cattle originated from the Near east domestication center, and introgressed through the North part of the continent and from there they would have dispersed east, west and south. Then, indicine cattle were introduced to Africa and *Bos indicus* genes were introgressed into native populations through absorbent crosses⁴⁸. Currently, an east–west gradient of Zebu influence in African native genes is observed.

Subsequent dispersal and crossbreed processes described above (founder group, migration and gene introgression) and natural and artificial selection would have shaped the *BoLA-DRB3* diversity in the current bovine populations. Accordingly, the *BoLA-DRB3* alleles detected in the Sudanese cattle were interspersed distributions along the allele NJ tree instead of grouped in specific clusters of the dendrogram, which is consistent with the ancient origin of the *BoLA-DRB3* alleles. Similar results have been reported in other native cattle breeds from different geographical regions^{21,22}.

Our Venn diagram illustrates the distribution of allelic diversity among different bovine groups, demonstrating that 14 *BoLA-DRB3* alleles were only detected in the Sudanese cattle breeds. Seven of these alleles corresponded to new variants described in this study (Table 1). Furthermore, a review of the IPD–MHC database showed that this group of Sudanese private alleles included seven other variants previously detected only in African breeds (Table S4).

Two *BoLA-DRB3* alleles, that were only previously reported in Creole cattle breeds^{21,37}, were identified in native Sudanese breeds. Studies based on mitochondrial DNA and Y chromosome haplotypes have revealed an African component in the germplasm of the American creole bovine breeds. Two origins have been proposed for this African component: through the native Iberian cattle that are the ancestors of Creole cattle and/or a direct introgression from mainland Africa following the slave trade routes⁴⁹. The Iberian theory is unlikely as the *BoLA-DRB3*011:02* and *BoLA-DRB3*029:02* alleles have not been detected in the Spanish Morucha breed, which were only autochthonous Iberian breed in which the genetic diversity of the *BoLA-DRB3* gene has been studied so far²⁰. In summary, 16 possible African putative alleles were identified in the native bovine populations of Sudan, totaling 20.22% of the gene frequency. The presence of private *BoLA-DRB3* alleles (not detected in zebu breeds so far) in native African breeds with humped phenotype suggest that current global diversity of this gene could have been retained in the founder group that originate African taurine native breeds⁴⁵.

On the other hand, a group of alleles is shared between the Sudanese breeds and the Zebu, Southeast Asian and/or Creole American breed groups (Table S5), but is absent in the European breeds. It is worth noting that these alleles were first identified in cattle breeds such as Boran, Ethiopian Arsi, N'Dama and Brahman (^{32,34,50}; <https://www.ebi.ac.uk/ipd/mhc/group/BoLA/>) (Table S5). The introgression of these variants could have been a consequence of the successive waves of introduction of Zebu cattle into the African continent⁴⁸. These alleles account for an additional 15.33% of the gene frequencies. The remaining alleles have a worldwide geographical distribution; thus, 20 variants have been detected in all the breed groups included in the Venn diagram. Further studies on the genetic diversity of the *BoLA-DRB3* gene in other African bovine populations will surely reveal a greater allelic repertoire.

The current repertoire of alleles of the *BoLA-DRB3* gene in the native cattle of Sudan would not only have been molded by stochastic forces, such as the formation of the founder group, gene drift and recent or historical gene introgression as described above, but also by processes of natural and artificial selection. In Sudan, as in other African regions, cattle are subjected to strong environmental pressures, such as tropical diseases, heat stress, drought and poor nutritional and forage deficits. Furthermore, animals are affected by diverse infectious diseases, including parasites (e.g., ticks, theileriosis, babesiosis, anaplasmosis, trypanosomosis;^{51–57}, bacteria (e.g., Hemorrhagic septicemia, Anthrax, tuberculosis, brucellosis, Thrombotic meningoencephalitis;^{58–62}) and viruses (e.g., foot and mouth disease, lumpy skin disease, Pox virus, bovine viral diarrheal diseases complex;^{53,63,65}). For this reason, it is to be expected that native Sudanese cattle will be under strong selection pressure, which would contribute to maintaining and shaping the genetic diversity of the *BoLA-DRB3* gene. In this sense, a wide repertoire of alleles allows the population to identify and respond to a greater range of antigens. Furthermore, heterozygous animals trigger an immune response to a greater variety of antigens. For these reasons, it has been proposed that this allelic diversity is maintained by balancing or over-dominant selection^{30,65,66}. Different indices at the population, nucleotide and amino acid levels showed high levels of genetic diversity in the bovine breeds of Sudan for the *BoLA-DRB3* gene. This is clearly reflected in the presence of a homogeneous distribution of gene frequencies (a high number of alleles with low frequencies). This is particularly extreme in the Baggara breed in which Slatkin's neutrality test showed evidence that the *BoLA-DRB3* gene frequency profile showed an even distribution consistent with the theoretical proportion expected under balancing selection pressures. Similar results have been reported for other cattle breeds, including Japanese Black, Yacumeño Creole, Bolivian Gir, Pyer Sein and Shwe Ni^{21,22,30}. Furthermore, the selection index (ω) revealed the presence of diversifying selection in several amino acid sites (mainly in the ABS) in *BoLA-DRB3* exon 2 of the Sudan native breeds. In contrast, the HWE test did not detect the effect of over-dominant selection⁶⁷. As discussed previously²¹, this effect has been observed only in some of the breeds studied so far and the most common explanation for the absence of heterozygote excess in the studied bovine breeds is the magnitude of the overdominance selection coefficient at MHC loci (probably lower than 0.02;⁶⁸). Such selection would only be enough to increase the number of heterozygotes in large populations and in the absence of high rates of stochastic forces (population bottlenecks,

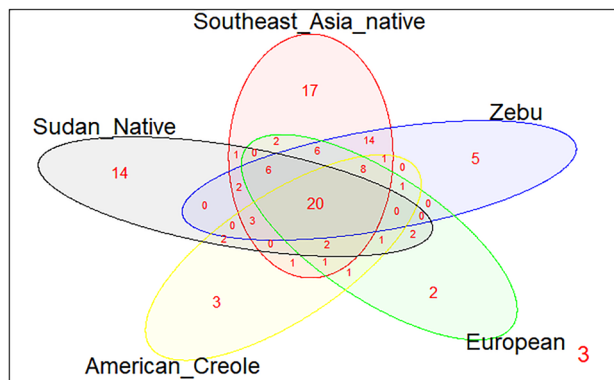


Figure 2. Venn plot of *BoLA-DRB3* alleles shared by Sudan native (Baggara, Kenana, and Butana); Southeast Asia (Myanmar and Philippine native breeds); indicine (Nellore, Gir, Brahman, and crossbreeds); European (Hereford, Black and Red Angus, Jersey, Shorthorn, Holstein, overo negro, overo colorado, and crossbreeds); and American Creole (Yacumeño and Hartón del Valle) cattle breeds.

genetic drift, and inbreeding). For this reason, and because the HWE method may suffer from low resolving power, such effects were not observed.

The repertoire of alleles of the *BoLA-DRB3* gene present in the native cattle of Sudan allows these breeds to be clearly differentiated from the rest, forming a cluster in the NJ trees and a narrow cloud in the PCA. This pattern is confirmed when PCAs are performed based on the pocket 4 and pocket 9 gene frequencies. It has previously been proposed that pocket 4 plays an important role in the binding of peptides due to this pocket being located in the center of the PBC^{69,70}. In addition, it has been reported in cattle that immune responses against vaccine and disease resistance is significantly related to differences in the pocket 4 motif^{49,50}. A particular amino acid (e.g., amino acid R in position 70) or amino acid motifs (e.g., ER at 70 and 71 sites; EIAY motif at positions 66–67–74–78, and the deletion of the amino acid 65), in sites that affect the conformation of pocket 4, have been associated with immune response or resistance to infectious diseases, such as mastitis, persistent lymphocytosis, dermatophilosis, and tick-borne diseases^{25,50,69,71–73}. Many of these diseases, as well as others mentioned above, are present in Sudan and could have contributed to shaping the current repertoire of *BoLA-DRB3* alleles present in native Sudanese cattle. However, these results were obtained in breeds that have different genetic backgrounds and that are raised in different environments and production systems, so further association studies are necessary to determine the effect (resistance or susceptibility) of the alleles present in the native cattle breeds of Sudan against different infectious diseases.

Conclusions and future prospects

To the best of our knowledge, this is the first study to document in detail the genetic diversity (taurine vs indicine) of *BoLA-DRB3* alleles in cattle not only in Sudan but in the entire African continent. In addition to the clear genetic clustering of cattle based on ancestral origin and phylogeography, we identify seven novel alleles in the three native Sudanese cattle breeds. Two evolutionary forces appear to contribute to the preservation and shaping of the genetic diversity of the *BoLA-DRB3* gene in native Sudanese cattle; diversifying selection mainly affects the ABS of the native breeds and balancing selection. The results demonstrate that the background variation between two cattle groups, taurine and indicine, is primarily due to events of origin, selection, and adaptation, which explains the variations found in the diversity of the *BoLA-DRB3* genes, not only between the two major groups but also with the indicine cattle group. This variation may explain how cattle from Sudan are resistant to various diseases. We presume that this genetic information provides a basis for better design of suitable breeding schemes. This variation may contribute to resistance in Sudanese cattle to various diseases.

Materials and methods

Sampled populations and genomic DNA extraction. The ODK (Open Data Kit) system was used to record the sampling information: breed name, sex, estimated age, sampling location GPS coordinates, photo of the animal and owner's information. All methods were carried out in accordance with relevant guidelines and regulations of the Faculty of Veterinary Medicine, University of Khartoum (Vet. Med. U of K), and all experimental protocols were approved by the Vet. Med. U of K research board committee. Before animals were sampled, written informed consents were obtained from all animal owners. Three cattle breeds were examined: (1) Butana breed: collected from the Atbara Butana Station and surrounding villages and from El-Gadarif city and Butana plain; (2) Kenana breed: samples were collected from Rabak city and surrounding villages and from UmBanein Kenana Station; (3) Baggara breed populations (i) Nyalawi population, which is a western Baggara breed sampled from calves from Nyala city, South Darfur; (ii) Daeinawi population, from Ed daein city. Whereas Nyalawi are large white cattle, some with black splashes, the Daeinawi are smaller and red with black along the neck and lateral sides of the head, hind quarters and shoulder sides (Fig. S4).

A total of 225 native breed cattle were sampled: Baggara N = 113, Butana N = 60 and Kenana N = 52 (Table S1 and Fig. S4). Seven milliliters of venous blood were collected in EDTA-containing vacutainer tubes. Genomic

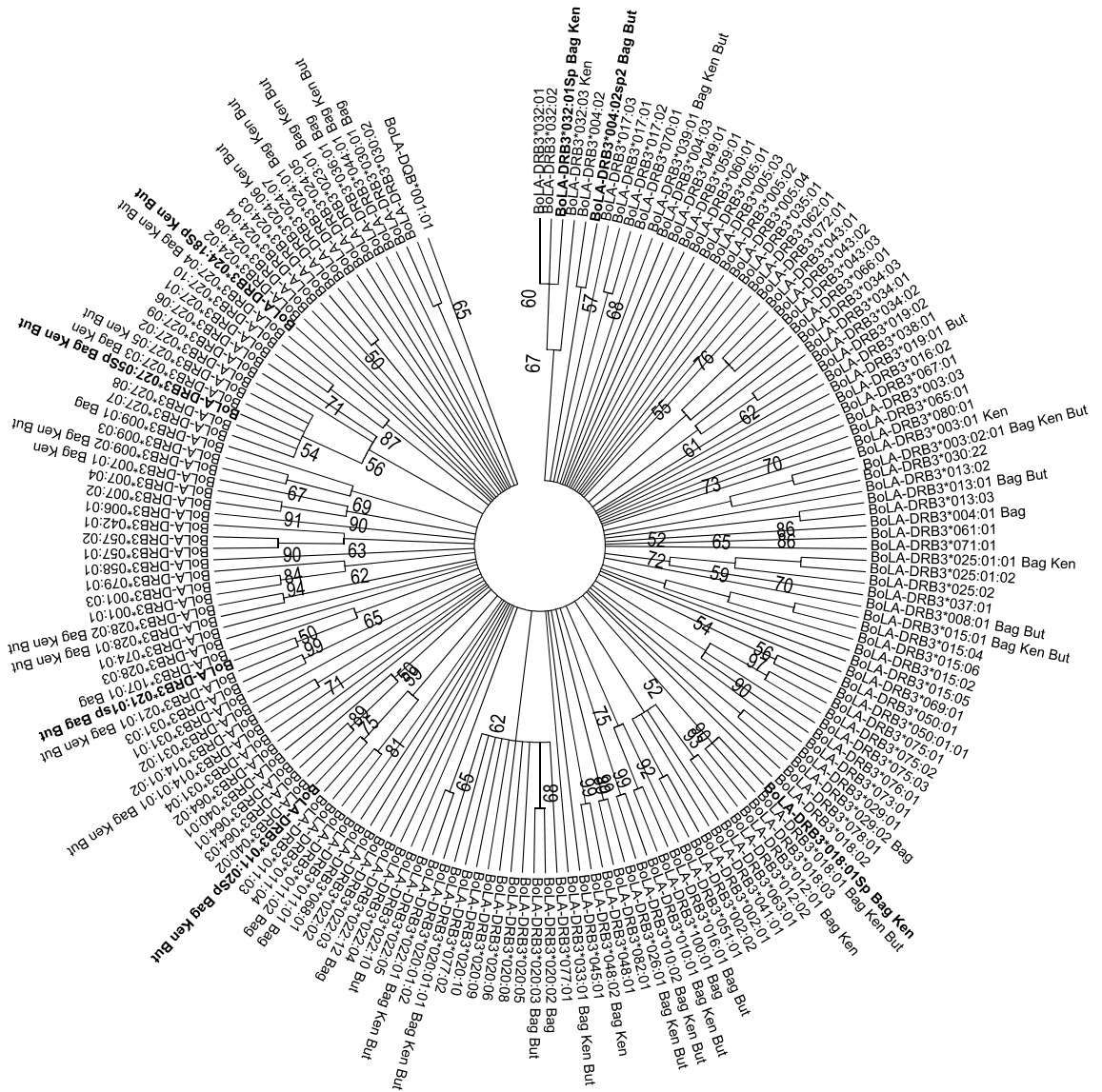


Figure 3. Neighbor-joining (NJ) tree constructed from the 270 bp nucleotide sequence that includes the β 1 domain encoded by all reported *BoLA-DRB3* alleles and the seven new ones (*BoLA-DRB3*004:02Sp2*, *BoLA-DRB3*011:02Sp*, *BoLA-DRB3*018:01Sp*, *BoLA-DRB3*021:01sp*, *BoLA-DRB3*024:18Sp*, *BoLA-DRB3*027:05sp*, and *BoLA-DRB3*032:01sp*). Numbers are bootstrap percentages that support each node. Bootstrapping was carried up with 1000 replicates to access the reliability of individual branches. Bag = Baggara, But = Butana, Ken = Kenana. Arrows indicate novel alleles.

Breed	π	NPD	Total			ABS		
			ds	dn	dn / ds	ds	dn	dn / ds
Baggara	0.080	19.36	0.030	0.105	3.50	0.090	0.435	4.83
Butana	0.074	17.99	0.028	0.105	3.75	0.093	0.447	4.80
Kenana	0.075	18.28	0.027	0.104	3.85	0.090	0.444	4.93

Table 3. Nucleotide diversity (π), mean number of pairwise differences (NPD) and mean number of non-synonymous (d_n) and synonymous (d_s) nucleotide substitutions per site.

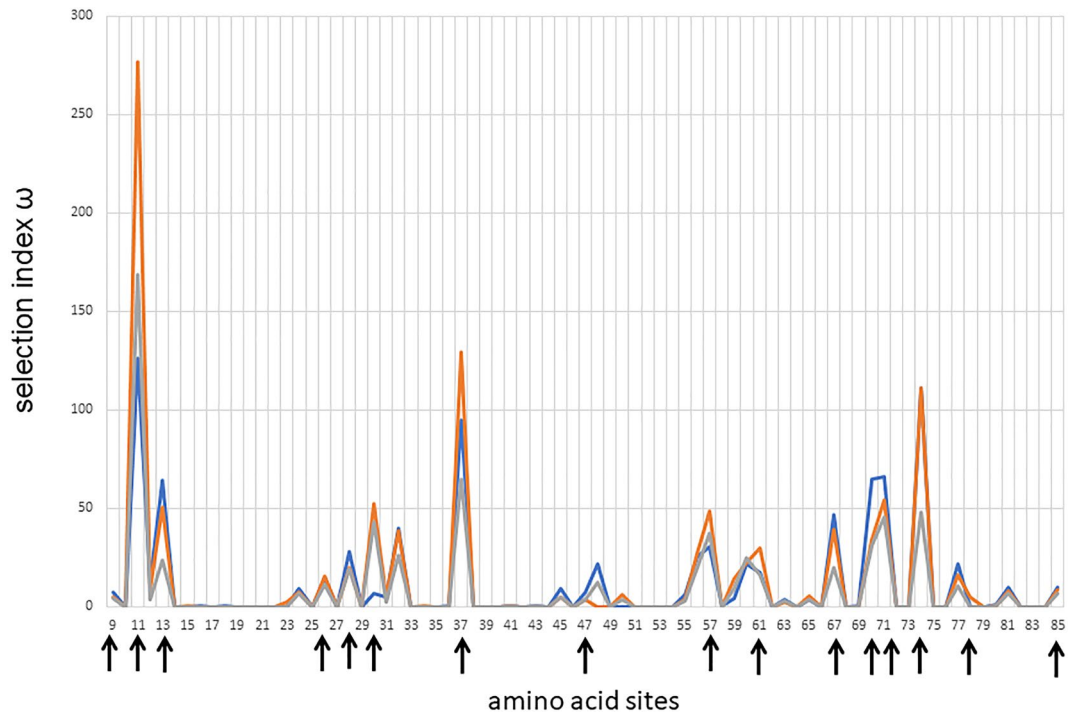


Figure 4. Estimated values of the selection index ω in each amino acid site along *BoLA-DRB3* exon 2 in Baggara (grey), Kenana (blue) and Butana (brown) Sudan cattle breeds. Arrows indicate the antigen-binding site (ABS).

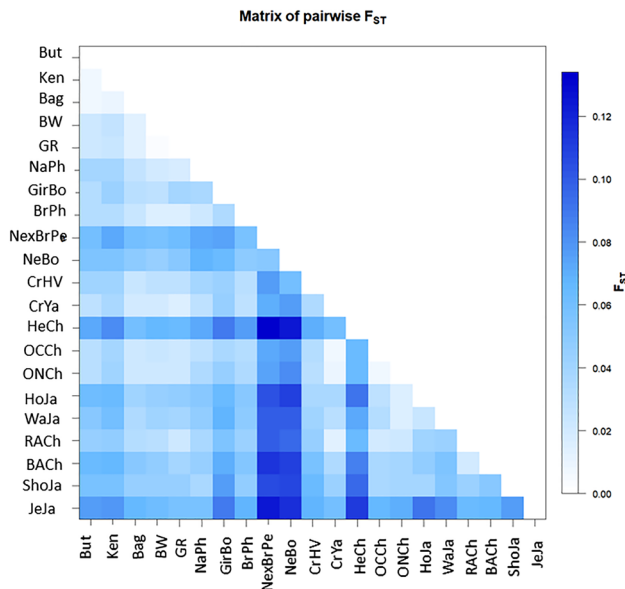


Figure 5. Graphic representation of calculated F_{ST} between population pairs using an R package pairFstMatrix.r. But = Butana, Ken = Kename, Bag = Baggara, BW = Pyer Sein. GR = Shwe Ni, NaPh = Philippine native, GirBo = Bolivian Gir, BrPh = Philippine Brahman, BrxNePe = Peruvian Brahman \times Nellore crossbreed, NeBo = Bolivian Nellore, CrHV = Creole Hatón del Valle, CrYa = Creole Yacumeño, HeCh = Chilean Hereford, OCCh = Chilean Overo Colorado, ONCh = Chilean Overo Negro, HoJa = Japanese Holstein, WaJa = Japanese Black, BACH = Chilean Black Angus, RACH = Chilean Red Angus, ShJa = Japanese Shorthorn and JeJa = Japanese Jersey.

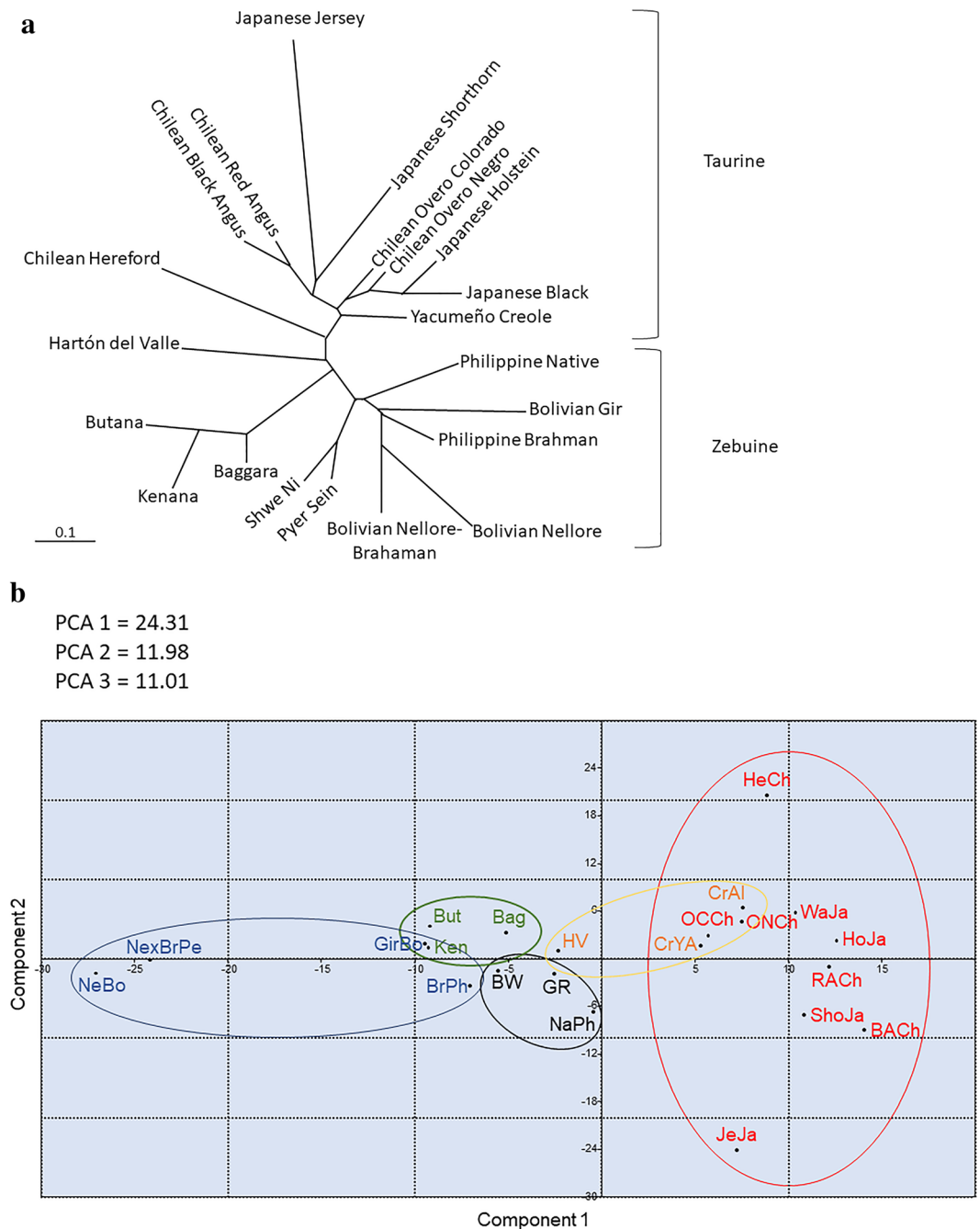


Figure 6. (a) Neighbor-joining dendrogram constructed from a matrix of D_A genetic distances. (b) Principal Component Analysis of allele frequencies from the *BoLA-DRB3* gene in 22 breeds. But = Butana, Ken = Kename, Bag = Baggara, BW = Pyer Sein. GR = Shwe Ni, NaPh = Philippine native, GirBo = Bolivian Gir, BrPh = Philippine Brahman, BrxNePe = Peruvian Brahman \times Nellore crossbreed, NeBo = Bolivian Nellore, CrHV = Creole Hatón del Valle, CrYa = Creole Yacumeño, Creole Highland, HeCh = Chilean Hereford, OCCh = Chilean Overo Colorado, ONCh = Chilean Overo Negro, HoJa = Japanese Holstein, Waja = Japanese Black, BACH = Chilean Black Angus, RACH = Chilean Red Angus, ShJa = Japanese Shorthorn and JeJa = Japanese Jersey.

DNA was extracted using DNeasy® Blood and Tissue Kit, (Qiagen, Germany), following the manufacturer's instructions.

PCR amplification and sequencing. Exon 2 of the *BoLA-DRB3* was amplified by PCR as described by²⁶. Using DRB3FRW 5-CGCTCCTGTGA(C/T)CAGATCTATCC-3 and DRB3REV 5-CACCCCCGCGCTCACC-3, PCR reactions were performed in a 25 μ l-reaction mixture containing 12.5 μ l of 2 \times Gflex PCR Buffer (Mg^{+2} , dNTP plus) (TaKaRa Bio Inc., Shiga, Japan), and 0.5 μ l of Tks Gflex DNA polymerase (1.25 units/ μ l) (TaKaRa

Sudan Breeds
 PCA 1 = 34.65
 PCA 2 = 30.66
 PCA 3 = 25.24

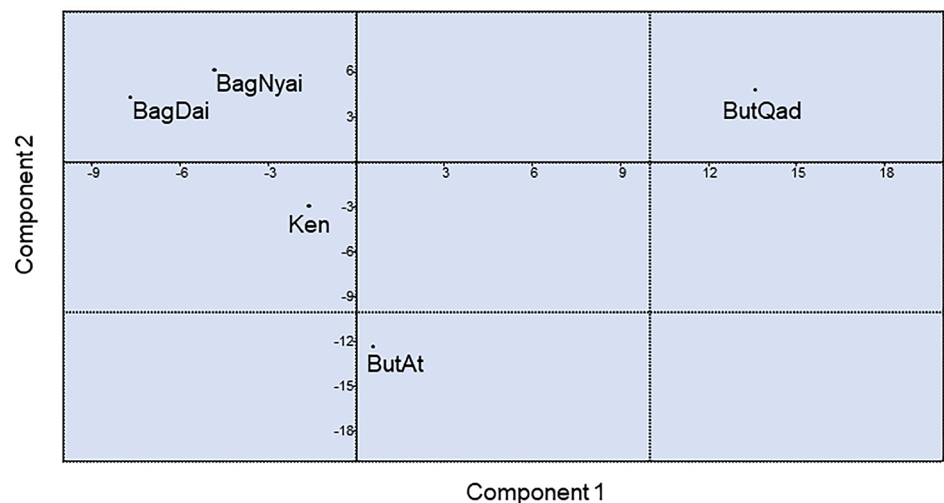


Figure 7. Principal components analysis of allele frequencies from the BoLA-DRB3 gene in five Sudan native samples sites (BagDai = Baggara Daiwani, BagNyai = Baggara Nyakawi, Ken = Kenana, ButAt = Butana Bu Atbara, and ButQad = Butana Bu Qadarif).

Bio Inc.), 200 nM of each primer, and 1.0 μ l of template. The reaction conditions consisted of an initial denaturation step at 95 $^{\circ}$ C for 3 min, followed by 35 cycles of 95 $^{\circ}$ C for 1 min, 58 $^{\circ}$ C for 30 s and 68 $^{\circ}$ C for 90 s and a final extension step at 68 $^{\circ}$ C for 5 min. PCR products were purified using a NucleoSpin Gel and PCR Clean Up Kit (Takara Bio Inc.). Cycle sequencing reactions were performed directly using the two PCR primers using the BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and analyzed on an ABI Prism 3130 \times genetic analyzer (Applied Biosystems) according to the manufacturer's instructions.

Sequence data analysis. Prior to analysis, all the chromatograms were visualized and sequence fragments were edited manually using ATGC software version 9.1 (GENETYX Corporation, Tokyo, Japan) correcting base calling errors. Multiple sequence alignments were performed using the MUSCLE algorithm implemented in MEGA X⁷⁴, and were subsequently joined to reconstruct a fragment of 280 bp spanning the entire exon 2.

BoLA-DRB3 allele genotyping. For typing *BoLA-DRB3* genotypes, we used the method implemented by²⁶. First, we downloaded a MHC_nuc.txt file from the IPD-MHC in order to update the allele database. This file contains all reported BoLA-DRB3 alleles. Then DNA sequences from the cattle for both strands (forward and reverse ab1 files) were imported together into the Assign 400ATF ver. 1.0.2.45 software (Conexio Genomics, Fremantle, Australia), which automatically aligned the sampled cattle sequences with those of previously reported *BoLA-DRB3* sequences, building a consensus. The most likely genotype is shown in the same window as the chromatograms so that they can be crosschecked. When we found a clear mismatch from several samples, we assigned these samples containing new alleles and revised the *BoLA-DRB3* database containing new allele sequences. The accuracy of the in silico genotyping method was demonstrated in Takeshima et al. (2001, 2011) where the new detected alleles were confirmed by cloning and sequencing, and the used method was developed and validated for only the *BoLA-DRB3* gene. If the sample could not genotype using these criteria, we discarded the sample result from this analysis.

Statistical analyses. *Genetic diversity at allele level.* Allele frequencies and the number of alleles (n_a) were obtained by direct counting. The distribution of alleles across breeds was analyzed by a Venn plot created using the R package 'VennDiagram' (<http://cran.r-project.org/>). The observed (h_o) and unbiased expected (h_e) heterozygosity of the *BoLA-DRB3* locus were estimated according to⁷³ using the Arlequin 3.5 software for population genetic analyses⁷⁶ (Schneider, 2000). F_{IS} statistics⁷⁷ for each breed were calculated using the Exact Test included in Genepop 4.7 software⁷⁸ to evaluate deviation from Hardy–Weinberg equilibrium (HWE). The Ewens–Watterson–Slatkin Exact Test of neutrality was carried out using the method described by⁷⁹ and implemented in the Arlequin 3.5 program.

Breed genetic structure. Genetic structure and genetic differentiation within Sudanese cattle breeds and among bovine breeds were assessed using Wright's F_{ST} statistics⁷⁷. This parameter was estimated using Arlequin 3.5 and Genepop 4.7 software. The F_{ST} values were represented graphically using the pairFstMatrix.r function implemented in the R statistical environment.

Genetic relationship between breeds. To condense the genetic variation at the *BoLA-DRB3* locus, allele frequencies were used to perform a PCA according to the⁸⁰ method, implemented in Past software⁸¹. Nei's standard genetic distances D_s ^{82,83} were calculated from allele frequencies and were used to perform cluster analysis using the Neighbor-Joining (NJ) algorithm⁸⁴. Confidence intervals for the groupings were estimated by bootstrap resampling of the data using 1000 replicates. Genetic distances and trees were computed using the Populations 1.2.28 software⁸⁴. The trees were then visualized using TreeView⁸⁵.

Genetic diversity at sequence level. Nucleotide diversity (π) and pairwise differences in nucleotide substitutions between alleles within each breed were calculated using Arlequin 3.5. The mean number of nonsynonymous (d_N), and synonymous (d_S) nucleotide substitutions per site from averaging over all sequence pairs were estimated within each group using the modified Nei-Gojobori model⁸³ and Jukes-Cantor's formula implemented in the software MEGA X⁷². The possibility that certain codon sites are under diversifying selection within each native Sudan breed was investigated using the Bayesian method implemented using OmegaMap⁸⁶. This method incorporates intragenic recombination and does not assume a known fixed genealogy, so that recombination does not inflate the false detection rate of positive sites⁸⁷. The *BoLA-DRB3* allele tree was constructed from a distance matrix that was based on the NJ method using the MEGA X software. Furthermore, a tree based only on ABS amino acid motifs was inferred using Maximum Parsimony method implemented in MEGA X. To test the significance of the branches of both trees, 1000 bootstrap replicate calculations were performed.

Data availability

Supplementary Material contains Table S1-S5 and Figures S1-S3 including detailed descriptions of all supplemental files.

Received: 26 August 2020; Accepted: 2 August 2021

Published online: 25 August 2021

References

- Mason, I. L. & Maule, J. P. The Indigenous Livestock of Eastern and Southern Africa. Farnham Royal (1960).
- Payne, W. J. A. *Cattle Production in the Tropics: General Introduction and Breeds and Breeding* (Longman, 1970).
- Bahbahani, H. *et al.* Signatures of positive selection in African Butana and Kenana dairy zebu cattle. *PLoS ONE* **13**, e0190446 (2018).
- Rege, J. E. O., Ayalew, W., Getahun, E., Hanotte, O. & Dessie, T. DAGRIS (Domestic Animal Genetic Resources Information System). International Livestock Research Institute, Addis Ababa, Ethiopia. Ababa, <http://dagris.ilri.cgiar.org> (2006).
- Salim, B., Taha, K. M., Hanotte, O. & Mwacharo, J. M. Historical demographic profiles and genetic variation of the East African Butana and Kenana indigenous dairy zebu cattle. *Anim. Genet.* **45**, 782–790 (2014).
- El-Habeeb, E. A. Variation in reproductive and milk production traits in Butana and Kenana dairy cattle in the Sudan. (PhD Thesis). University of Khartoum-Sudan (1991).
- Medani, M. A. *Animal Resources and Animal Production in Sudan* 2nd edn. (University of Khartoum Printing Press, 2003).
- Musa, L. M., Ahmed, M., Peters, K. J., Zumbach, B. & Gubartalla, K. E. The reproductive and milk performance merit of Butana cattle in Sudan. *Archiv fur Tierzucht.* **48**, 445 (2005).
- Alsiddig, M. A., Babiker, S. A., Galal, M. Y. & Mohammed, A. M. Phenotypic characterization of Sudan Zebu cattle (Baggara type). *Res. J. Anim. Vet. Sci.* **5**, 10–17 (2010).
- Cooper, M. D. & Alder, M. N. The evolution of adaptive immune systems. *Cell* **124**, 815–822 (2006).
- Klein, J. *Natural History of the Major Histocompatibility Complex*, 99th edition. ed. Wiley (1986).
- Kelley, J., Walter, L. & Trowsdale, J. Comparative genomics of major histocompatibility complexes. *Immunogenetics* **56**, 683–695 (2005).
- Ujvari, B. & Belov, K. Major histocompatibility complex (MHC) markers in conservation biology. *Int. J. Mol. Sci.* **12**, 5168–5186 (2011).
- Fries, R., Hediger, R. & Stranzinger, G. Tentative chromosomal localization of the bovine major histocompatibility complex by in situ hybridization. *Anim. Genet.* **17**, 287–294 (1986).
- Fries, R., Eggen, A. & Womack, J. E. The bovine genome map. *Mamm. Genome* **4**, 405–428 (1993).
- Kim, J. *et al.* The genome landscape of indigenous African cattle. *Genome Biol.* **18**(1), 34 (2017).
- Takeshima, S.-N., Ikegami, M., Morita, M., Nakai, Y. & Aida, Y. Identification of new cattle BoLA-DRB3 alleles by sequence-based typing. *Immunogenetics* **53**(1), 74–81 (2001).
- Takeshima, S.-N. *et al.* Characterization of bovine MHC class II DRB3 diversity in South American Holstein cattle populations. *Tissue Antigens* **86**, 419–430 (2015).
- Takeshima, S.-N., Saitou, N., Morita, M., Inoko, H. & Aida, Y. The diversity of bovine MHC class II DRB3 genes in Japanese Black, Japanese Shorthorn, Jersey and Holstein cattle in Japan. *Gene* **316**, 111–118 (2003).
- Bohórquez, M. D. *et al.* Major histocompatibility complex class II (DRB3) genetic diversity in Spanish Morucha and Colombian Normande cattle compared to taurine and zebu populations. *Front. Genet.* **10**, 1293 (2020).
- Giovambattista, G. *et al.* Characterization of bovine MHC DRB3 diversity in Latin American Creole cattle breeds. *Gene* **519**(1), 150–158 (2013).
- Giovambattista G. *et al.* Characterization of bovine MHC DRB3 diversity in worldwide cattle breeds focusing on Myanmar cattle. 37th Conference of the International Society of Animal Genetics. del 7 al 12 de julio del. Lleira, España (2019).
- Miyasaka, T., Takeshima, S.-N., Sentsui, H. & Aida, Y. Identification and diversity of bovine major histocompatibility complex class II haplotypes in Japanese Black and Holstein cattle in Japan. *J. Dairy Sci.* **95**, 420–431 (2012).
- Takeshima, S.-N., Matsumoto, Y. & Aida, Y. Establishment of a new polymerase chain reaction–sequence-based typing method for genotyping cattle major histocompatibility complex class II DRB3. *J. Dairy Sci.* **92**, 2965–2970 (2009).

25. Takeshima, S., Sarai, Y., Saitou, N. & Aida, Y. MHC class II DR classification based on antigen-binding groove natural selection. *Biochem. Biophys. Res. Commun.* **385**, 137–142 (2009).
26. Takeshima, S.-N.M. *et al.* A new method for typing bovine major histocompatibility complex class II DRB3 alleles by combining two established PCR sequence-based techniques. *Tissue Antigens* **78**, 208–213 (2011).
27. Takeshima, S.-N. *et al.* The great diversity of major histocompatibility complex class II genes in Philippine native cattle. *Meta Gene* **2**, 176–190 (2014).
28. Takeshima, S.-N., Nakai, Y., Ohta, M. & Aida, Y. Characterization of DRB3 alleles in the MHC of Japanese Shorthorn Cattle by polymerase chain reaction-sequence-based typing. *J. Dairy Sci.* **85**, 1630–1632 (2002).
29. Takeshima, S.-N. *et al.* Assessment of biodiversity in Chilean cattle using the distribution of major histocompatibility complex class II BoLA-DRB3 allele. *Tissue Antigens* **85**, 35–44 (2015).
30. Takeshima, S., Corbi-Botto, C., Giovambattista, G. & Aida, Y. Genetic diversity of BoLA-DRB3 in South American Zebu cattle populations. *BMC Genet.* **19**, 33 (2018).
31. Lee, B.-Y., Hur, T., Jung, Y.-H. & Kim, H. Identification of BoLA-DRB3.2 alleles in Korean native cattle (Hanwoo) and Holstein populations using a next generation sequencer. *Anim. Genet.* **43**, 438–441 (2012).
32. Gelhaus, A., Schnittger, L., Mehlitz, D., Horstmann, R. D. & Meyer, C. G. Sequence and PCR-RFLP analysis of 14 novel BoLA-DRB3 alleles. *Anim. Genet.* **26**, 147–153 (1995).
33. Maccari, G. *et al.* IPD-MHC 2.0: An improved inter-species database for the study of the major histocompatibility complex. *Nucleic Acids Res.* **45**, 860–864 (2017).
34. Mikko, S. & Anderson, L. Extensive MHC class II DRB3 diversity in African and European cattle. *Immunogenetics* **42**, 408–403 (1995).
35. Miyasaka, T. *et al.* The diversity of bovine MHC class II DRB3 and DQA1 alleles in different herds of Japanese Black and Holstein cattle in Japan. *Gene* **472**, 42–49 (2011).
36. Jakobsson, M., Edge, M. D. & Rosenberg, N. A. The relationship between F_{ST} and the frequency of the most frequent allele. *Genetics* **193**, 515–528 (2013).
37. Giovambattista, G. *et al.* BoLA-DRB3 genetic diversity in Highland Creole cattle from Bolivia. *HLA* **96**, 688–696 (2020).
38. Kemp, S. J., Spooner, R. L. & Teale, A. J. A comparative study of major histocompatibility complex antigens in East African and European cattle breeds. *Anim. Genet.* **19**, 17–29 (1988).
39. Klein, J., Satta, Y., Takahata, N. & O'hUigin, C. Trans-specific Mhc polymorphism and the origin of species in primates. *J. Med. Primatol.* **22**, 57–64 (1993).
40. Bradley, D. G., MacHugh, D. E., Cunningham, P. & Loftus, R. T. Mitochondrial diversity and the origins of African and European cattle. *PNAS* **93**, 5131–5135 (1996).
41. Loftus, R. T. *et al.* Mitochondrial genetic variation in European, African and Indian cattle populations. *Anim. Genet.* **25**, 265–271 (1994).
42. Loftus, R. T. *et al.* Evidence for two independent domestications of cattle. *Proc. Natl. Acad. Sci.* **91**, 2757–2761 (1994).
43. Troy, C. S. *et al.* Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**, 1088–1091 (2001).
44. Mannen, H. *et al.* Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle. *Mol. Phylogenet. Evol.* **32**, 539–544 (2004).
45. Mwai, O., Hanotte, O., Kwon, Y. J. & Cho, S. African indigenous cattle: unique genetic resources in a rapidly changing World. *Asian-Australas J Anim Sci.* **28**, 911–921 (2015).
46. Kim, K. *et al.* The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nat Genet.* **52**, 1099–1110 (2020).
47. Magee, D. A., MacHugh, D. E. & Edwards, C. J. Interrogation of modern and ancient genomes reveals the complex domestic history of cattle. *Anim. Front.* **4**, 7–22 (2014).
48. Hanotte, O. *et al.* African Pastoralism: genetic imprints of origins and migrations. *Science* **296**, 336–339 (2002).
49. Ginja, C. *et al.* The genetic ancestry of American Creole cattle inferred from uniparental and autosomal genetic markers. *Sci. Rep.* **9**, 11486 (2019).
50. Maillard, J.-C., Renard, C., Chardon, P., Chantal, I. & Bensaid, A. Characterization of 18 new BoLA-DRB3 alleles. *Anim. Genet.* **30**, 200–203 (1999).
51. Awad, H. *et al.* Prevalence and genetic diversity of Babesia and Anaplasma species in cattle in Sudan. *Vet. Parasitol.* **181**, 146–152 (2011).
52. Bakheit, M. A. & Latif, A. A. The Innate Resistance of Kenana Cattle to Tropical Theileriosis (*Theileria annulata* Infection) in the Sudan. *Ann. N. Y. Acad. Sci.* **969**, 159–163 (2002).
53. Elhassan, A. M., Fadol, M. A. & El-Hussein, A. M. Seroprevalence of bovine herpes virus-1, bovine herpes virus-4 and bovine viral diarrhoea virus in dairy cattle in Sudan. *Pak Vet J.* **31**, 317–320 (2011).
54. El Hussein, A. M., Majid, A. M. & Hassan, S. M. The present status of tick-borne diseases in the Sudan. *Arch Inst Pasteur Tunis.* **81**, 31–34 (2004).
55. Jongejan, F., Salih, D. A., El Hussein, A. M. & Hassan, S. M. Preliminary survey of ticks (Acari: Ixodidae) on cattle in northern Sudan. *Onderstepoort J. Vet. Res.* **71**, 319–326 (2004).
56. Salih, D. A. *et al.* Seroprevalence of tick-borne diseases among cattle in the Sudan. *Parasitol. Res.* **104**, 845–850 (2009).
57. Salim, B. *et al.* An outbreak of bovine trypanosomiasis in the Blue Nile State, Sudan. *Parasites Vectors* **4**, 74 (2011).
58. Abdalla, A. & Hamid, M. E. Comparison of conventional and non-conventional techniques for the diagnosis of bovine brucellosis in Sudan. *Trop. Anim. Health Prod.* **44**, 1151–1155 (2012).
59. Abusalab, S. M., Ahmed, A. M. & Hamid, M. E. Haemorrhagic septicaemia: a general review. *Sudan J Vet Res.* **18**, 1–14 (2003).
60. Ahmed, A. M., Mukhtar, M. M., ElHussein, A. M., Nourand, T. A. & Fadol, M. A. Combined anthrax and sheep pox vaccine, production and immunization trial in Sudan. *J. Anim. Vet. Adv.* **6**, 517–521 (2007).
61. Eltoum, K., Aradaib, I. & El-Sanousi, S. PCR-based randomly amplified polymorphic DNA (RAPD) fingerprinting for detection of genetic diversity among Sudanese isolates of *Haemophilus somnus*. *Veterinarski Arhiv.* **73**, 315–321 (2003).
62. Sulieman, M. S. & Hamid, M. E. Identification of Acid Fast Bacteria from caseous lesions in cattle in Sudan. *J. Vet. Med. Ser. B* **49**, 415–418 (2002).
63. Elsaïm, NAG-E. Prevalence of lumpy skin disease in cattle, in red sea state, Sudan. (MSc thesis). University of Khartoum, Sudan (2007).
64. Habiela, M. *et al.* Molecular Characterization of Foot-and-Mouth Disease Viruses Collected from Sudan. *Transbound. Emerg. Dis.* **57**, 305–314 (2010).
65. Hughes, A. L. & Nei, M. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *PNAS* **86**, 958–962 (1989).
66. Hughes, A. L. & Yeager, M. Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* **32**, 415–435 (1998).
67. Takeshima, S. *et al.* Evidence for cattle major histocompatibility complex (BoLA) class II DQA1 gene heterozygote advantage against clinical mastitis caused by Streptococci and Escherichia species. *Tissue Antigens* **72**, 525–531 (2008).
68. Mikko, S., Roed, K., Schmutz, S. & Andersson, L. Monomorphism and polymorphism at Mhc DRB loci in domestic and wild ruminants. *Immunol. Rev.* **167**, 169–178 (1999).

69. Baxter, R. *et al.* BoLA-DR peptide binding pockets are fundamental for foot-and-mouth disease virus vaccine design in cattle. *Vaccine*. **28**, 28–37 (2009).
70. Sharif, S., Mallard, B. A. & Sargeant, J. M. Presence of glutamine at position 74 of pocket 4 in the BoLA-DR antigen binding groove is associated with occurrence of clinical mastitis caused by *Staphylococcus* species. *Vet. Immunol. Immunopathol.* **76**, 231–238 (2000).
71. Garcia-Briones, M. M. *et al.* Association of bovine DRB3 alleles with immune response to FMDV peptides and protection against viral challenge. *Vaccine*. **19**, 1167–1171 (2000).
72. Sitte, K., Brinkworth, R., East, I. J. & Jazwinska, E. C. A single amino acid deletion in the antigen binding site of BoLA-DRB3 is predicted to affect peptide binding. *Vet. Immunol. Immunopathol.* **85**, 129–135 (2002).
73. Xu, A., Van Eijk, M. J., Park, C. & Lewin, H. A. Polymorphism in BoLA-DRB3 exon 2 correlates with resistance to persistent lymphocytosis caused by bovine leukemia virus. *J. Immunol.* **151**, 6977–6985 (1993).
74. Kumar, S., Nei, M., Dudley, J. & Tamura, K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **9**, 299–306 (2008).
75. Nei, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–590 (1978).
76. Schneider, S., Roessli, D. & Excoffier, L. Arlequin version 2.000: a software for population genetics data analysis. In: 2.000 edn: Genetics and Biometry Laboratory. Switzerland: University of Geneva (2000).
77. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
78. Rousset, F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106 (2008).
79. Slatkin, M. A correction to the exact test based on the Ewens sampling distribution. *Genet. Res.* **68**(3), 259–260 (1996).
80. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. The History and Geography of Human Genes. Princeton University Press (1994).
81. Hammer, Ø., Harper, David, A.T. & Ryan, P. D. PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*. **4**, 9 (2001).
82. Nei, M. Genetic distance between populations. *Am. Nat.* **106**, 283–292 (1972).
83. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* **3**, 418–426 (1986).
84. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
85. Langella O. Populations: a population genetic software. In: vol. CNRS UPR9034., 1.2.28 edn; 1999. <http://www.mybiosoftware.com/populations-12-32-population-genetic-software.html>.
86. Page, R. D. M. TreeView: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**, 357–358 (1996).
87. Wilson, D. J. & McVean, G. Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* **172**, 1411–1425 (2006).

Acknowledgements

We are grateful to the animal owners and Atbara Butana Station, UmBanein Kenana Station, Kuku Research Center and Multi-Purposes Company for allowing us to sample their animals. This study was supported by a grant for Young Scientist, Ministry of Higher Education, Sudan and Scientific Research, Sudan awarded to Dr Bashir Salim and partially supported by JSPS KAKENHI, Japan; grant numbers 16K21431 and 19H03118 to Dr. Ryo Nakao.

Author contributions

B.S. and J.M.M. conceived the initial study, B.S. designed all the described experiments. B.S., S.-N.T., G.G., analyzed the data and drafted the manuscript. M.-K.A., S.K. and A.M.A. helped in drafting the manuscript. M.A.-M.M. cleaned the sequence data. R.N. provided the laboratory facility, reagents and sequence analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-96330-7>.

Correspondence and requests for materials should be addressed to B.S.

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021