

RESEARCH ARTICLE

Open Access

The patterns of admixture, divergence, and ancestry of African cattle populations determined from genome-wide SNP data



N. Z. Gebrehiwot^{1*}, E. M. Strucken¹, H. Aliloo¹, K. Marshall² and J. P. Gibson^{1*}

Abstract

Background: Humpless *Bos taurus* cattle are one of the earliest domestic cattle in Africa, followed by the arrival of humped *Bos indicus* cattle. The diverse indigenous cattle breeds of Africa are derived from these migrations, with most appearing to be hybrids between *Bos taurus* and *Bos indicus*. The present study examines the patterns of admixture, diversity, and relationships among African cattle breeds.

Methods: Data for ~40 k SNPs was obtained from previous projects for 4089 animals representing 35 African indigenous, 6 European *Bos taurus*, 4 *Bos indicus*, and 5 African crossbred cattle populations. Genetic diversity and population structure were assessed using principal component analyses (PCA), admixture analyses, and Wright's *F* statistic. The linkage disequilibrium and effective population size (*N_e*) were estimated for the pure cattle populations.

Results: The first two principal components differentiated *Bos indicus* from European *Bos taurus*, and African *Bos taurus* from other breeds. PCA and admixture analyses showed that, except for recently admixed cattle, all indigenous breeds are either pure African *Bos taurus* or admixtures of African *Bos taurus* and *Bos indicus*. The African zebu breeds had highest proportions of *Bos indicus* ancestry ranging from 70 to 90% or 60 to 75%, depending on the admixture model. Other indigenous breeds that were not 100% African *Bos taurus*, ranged from 42 to 70% or 23 to 61% *Bos indicus* ancestry. The African *Bos taurus* populations showed substantial genetic diversity, and other indigenous breeds show evidence of having more than one African taurine ancestor. *N_e* estimates based on *r*² and *r*²_{adj} showed a decline in *N_e* from a large population at 2000 generations ago, which is surprising for the indigenous breeds given the expected increase in cattle populations over that period and the lack of structured breeding programs.

Conclusion: African indigenous cattle breeds have a large genetic diversity and are either pure African *Bos taurus* or admixtures of African *Bos taurus* and *Bos indicus*. This provides a rich resource of potentially valuable genetic variation, particularly for adaptation traits, and to support conservation programs. It also provides challenges for the development of genomic assays and tools for use in African populations.

Keywords: Admixture, African crossbreds, African indigenous, *Bos taurus*, *Bos indicus*, Effective population size, Genetic differentiation, Linkage disequilibrium, SNPs

* Correspondence: bgebrehi@myune.edu.au; jgibson5@une.edu.au

¹Centre for Genetic Analysis and Applications, School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia
Full list of author information is available at the end of the article



© The Author(s). 2020 **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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Based on skeletal evidence, Sahara rock art, and Egyptian dynastic representations, the humpless taurine cattle (*Bos taurus*) are thought to be the earliest domestic cattle in Africa [1]. Archaeological evidence suggested that African cattle were domesticated in the eastern Sahara 10,000 to 8000 years before present (BP) by hunter-gatherers [2]. But genetic evidence suggests a single domestication event in the Near East and subsequent crossing with wild aurochs in the southern Fertile Crescent and/or North Africa [3]. Using genome-wide SNP data of 67 ancient Near Eastern *Bos taurus* and modern populations, Verdugo et al. [4] suggested that the ancient Levantine genome affinity with Moroccan aurochs implies that the distinct phenotypes and genotypes in African *Bos taurus* cattle may stem from roots in the southern Fertile Crescent. In their review of the evidence, Stock and Gifford-Gonzalez [5] concluded that *Bos taurus* cattle likely spread across the Sinai and into the Nile Delta 7000 to 8000 BP, then across North Africa, and subsequently into the Nile and the then-grassy Sahara, possibly with additional inputs through the Horn of Africa. Humpless, presumably *Bos taurus* cattle first appear to be present south of the Sahara about 4500 to 4000 BP [6, 7]. A few depictions of Egyptian cattle show humped animals, which are claimed as evidence for the presence of *Bos indicus* cattle in Egypt from 3500 BP [8]. The earliest evidence for *Bos indicus* cattle in sub-Saharan Africa is in East Africa, where all samples, that could be analyzed from two sites dated around 2000 to 2500 BP, were of *Bos indicus* or Sanga (a hybrid of *Bos indicus* and *Bos taurus*) type [9]. This suggests that *Bos indicus* genes were already predominant in the pastoral systems in this region. Payne and Hodges [6] concluded that *Bos taurus* cattle, however, remained predominant in Ethiopia and East Africa until recently despite many waves of *Bos indicus* introductions to the region from about 2500 BP onwards.

Currently, Africa is home to more than 180 cattle breeds or distinct cattle populations [10], and several authors have made classifications of present-day African indigenous breeds of cattle. Rege and Tawah [11] suggested four categories of indigenous breeds: *Bos taurus*, *Bos indicus* (zebu), Sanga (*Bos taurus* × *Bos indicus* hybrid), and Zenga (Sanga×zebu hybrid). According to Lenstra and Bradley [12], African *Bos taurus* breeds are those that have short ears and no hump, while zebu breeds are those that have long floppy ears and a prominent hump. Subsequent results based on molecular marker data [13] and results presented here show that the genetic diversity of African cattle is more complex than this, most particularly, no African indigenous breeds have been shown to be pure *Bos indicus*. Thus, the term “zebu”, as applied to African cattle breeds,

means that the breed has a hump, but it does not imply that the breed is pure *Bos indicus*, despite much of the literature using zebu and *Bos indicus* as synonymous when applied to African cattle.

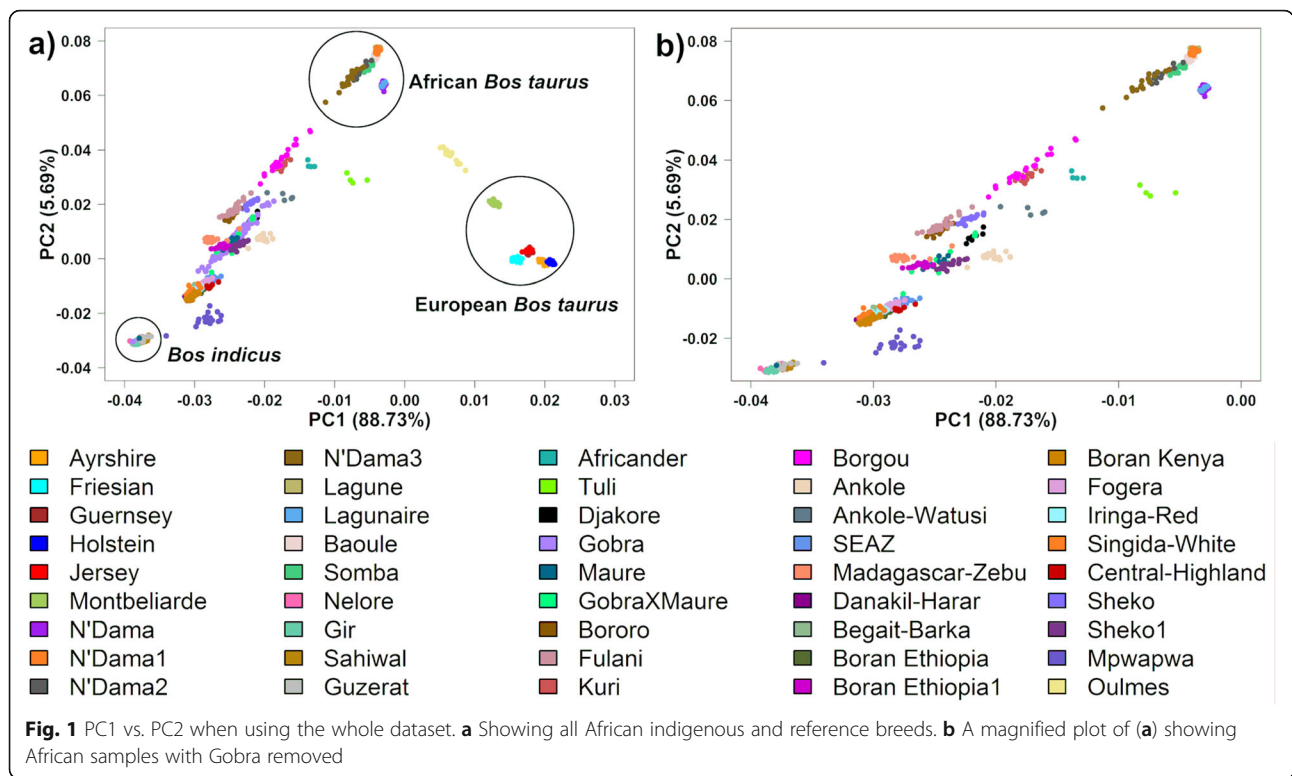
Studies of mitochondrial DNA (mtDNA) variation indicated that the two major groups of cattle, *Bos taurus* and *Bos indicus*, were genetically distinct before domestication [14–16]. A PCA result by Verdugo et al. [4] using genome sequence data on ancient cattle samples revealed that cattle origins consisted of two divergent aurochs populations that formed the basis of the *Bos indicus* and *Bos taurus* divide. These authors also showed, using mtDNA sequence data, that there was male-driven *Bos indicus* introgression into the Near East *Bos taurus* populations. Studies of microsatellite DNA and Y-chromosomal markers showed extensive introgression of male *Bos indicus* genes into existing African cattle populations [17–19], all of which currently carry *Bos taurus* mtDNA, indicating male-driven introgression of *Bos indicus* genes into the previously *Bos taurus* African cattle populations. Based on genome-wide autosomal SNP markers, Weerasinghe et al. [13] showed that all indigenous cattle breeds from Tanzania, Kenya, Uganda, and Ethiopia were admixtures of *Bos indicus* and African *Bos taurus*.

The present study provides one of the most extensive analyses of the genetic diversity of African cattle breeds based on genome-wide SNP data to date. We undertook admixture and principal component analyses, Wright’s F statistic (F_{ST} and F_{IS}), and linkage disequilibrium (LD) analyses to obtain a clear picture of the patterns of admixture and genetic diversity of African indigenous and crossbred populations and to compare their diversity to exotic breeds.

Results

Principal components and admixture analyses of indigenous breeds

Principal component analyses were performed to explore and visualize the genetic variation between different breeds and to identify potential sub-structures within the data. The first five principal components (PC) obtained from an analysis of all indigenous and crossbred cattle populations from East and West Africa, and including African and European taurine reference breeds as well as indicine reference breeds, explained a total of 96.1% of the variation in the genomic relationship matrix (GRM). The first two components accounted for 88.7 and 5.7% of the total genetic variation, respectively, and differentiated the *Bos indicus*, European *Bos taurus*, and African *Bos taurus* breeds from each other as the apexes of a triangle in the plot area (Fig. 1a). The indicine reference breeds, Nelore, Gir, Sahiwal, and Guzerat, grouped tightly together while the African taurine populations



clustered in two distinct groups (Figs. 1a, b, and S1a). The first African taurine group comprised N'Dama (from Guinea) and Lagunaire, and the second group included N'Dama1 (from Cote d'Ivoire), N'Dama2 (from Southeast Burkina Faso), N'Dama3 (from Southwest Burkina Faso), Lagune, Baoule and Somba. N'Dama2 and especially N'Dama3 appeared to include animals that spread towards the pooled *Bos indicus* reference breeds, showing that they are not pure African taurine breeds (Fig. 1a, b), and, therefore, these breeds were excluded from the African taurine reference breeds in later Admixture analyses.

A separate PCA was performed to evaluate in more detail the genetic structure among the eight African taurine reference populations (Figure S2). The first, second, and third PCs explained 32.1, 20.5, and 7% of the total variance among the African taurine breeds, respectively. Somba and Baoule clustered tightly together, while all other samples formed separate single clusters, except N'Dama3, which split into two clusters (for more detailed results, see Gebrehiwot et al. [20]).

The majority of the East African indigenous breeds that are classified as zebu breeds (Danakil-Harar, Begait-Barka, Ethiopian Boran, Fogera, Iringa-Red, Singida-White, Kenyan Boran, Central Highland, and SEAZ), clustered together on or slightly to the right and at the indicine end of the axis between indicine and the first

African taurine group (N'Dama and Lagunaire, Fig. 1b). Note that in Fig. 1b, the Gobra sample has been removed as it is not a pure breed sample, and it obscured the position of other samples in the plot. The Sheko1, Maure, Boran Ethiopia1, and Madagascar-zebu clustered further towards the Africa taurine breeds (i.e., lower *Bos indicus* admixture) and spread between the two axes that connect the indicine to the first African taurine group (axis 1) versus the second African taurine group (axis 2). Most of the hybrid animals between Gobra and Maure (Gobra x Maure) sit in this second cluster, aligning with axis 1. The Madagascar-zebu is distinct from all the zebu breeds being the only zebu breed that sits on axis 2.

The Ankole, Djakore, and Sheko (Sanga breeds), and Bororo and Fulani (zebu breeds) form the third cluster located more towards the African taurine breeds, with Ankole and Djakore close to axis 1 and the other breeds on or slightly to the left of axis 2 (Fig. 1b). The Bororo (also known as Red Fulani) and Fulani clustered together. Gobra showed a large genetic diversity along axis 1 (Fig. 1a). The Borgou and Kuri lie on axis 2, and the Ankole-Watusi and Africander lie on axis 1, all more towards African taurine than other breeds. The Tuli forms an outlier group consistent with high African taurine ancestry but well to the right of axis 1 indicating admixture with European taurine. Except for one outlier, the composite dual-purpose Mpwapwa breed clustered at the indicine end but well to the right of axis 1. The

Moroccan Oulmes Zaer clustered in an intermediary position between African and European taurine breeds.

Figures 2 and 3 illustrate the estimated breed ancestries from supervised Admixture Models 1 and 2 with $K=7$ and $K=11$, respectively. In Model 1, only one African taurine breed (N'Dama) was used together with a pooled indicine sample and five European taurine reference breeds. Model 2 included an additional four African taurine reference breeds to differentiate the African *Bos taurus* background. Consistent with the PCA, all African indigenous breeds, other than the pure African taurine breeds, were estimated to be an admixture of indicine and African taurine ancestries. Some breeds also showed small admixture with European taurine. Absolute estimates of ancestral proportions differed substantially between Admixture Model 1 versus Model 2, with Model 2 giving lower estimates of indicine ancestry. However, the ranking of breeds for indicine ancestry proportion was very similar between Model 1 and Model 2, and the following results summarised here are for Model 1. Overall, the indicine proportion was lower in West and South African breeds compared to East African breeds. However, the West African breeds, especially from Senegal, showed a wide range of *Bos indicus* ancestry. For example, the indicine component in Gobra ranged from 48.5 to 79.8% (average 65.3%), from 64.8 to 70.3% (average 67.8%) in Maure, and from 56.0 to 77.3% (average 66.2%) in Gobra x Maure crosses (Table 1).

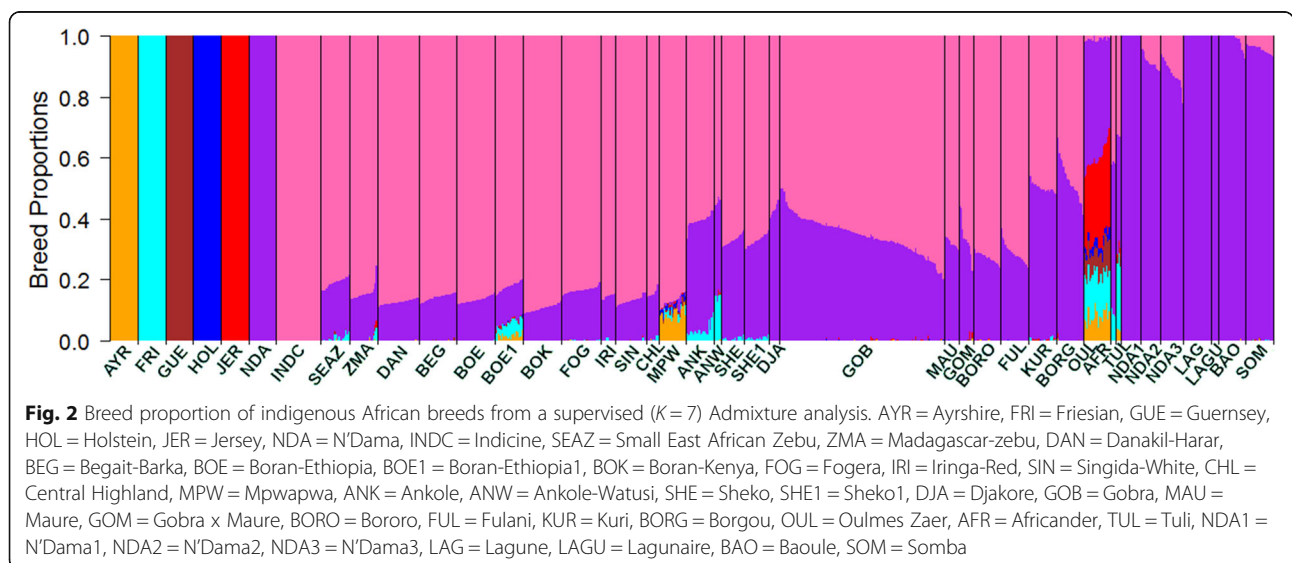
In East Africa, the Ankole, Ankole-Watusi, Sheko, and Sheko1 showed the lowest indicine proportions ranging from 55.3 to 67.7% (Fig. 2 and Table 1). Ankole-Watusi, Ankole, and Ethiopian Boran1 showed average exotic breed proportions larger than 1%. Ankole-Watusi had 13% exotic taurine ancestry, which was attributed mainly to Friesians and Ayrshires based on Model 1. The South

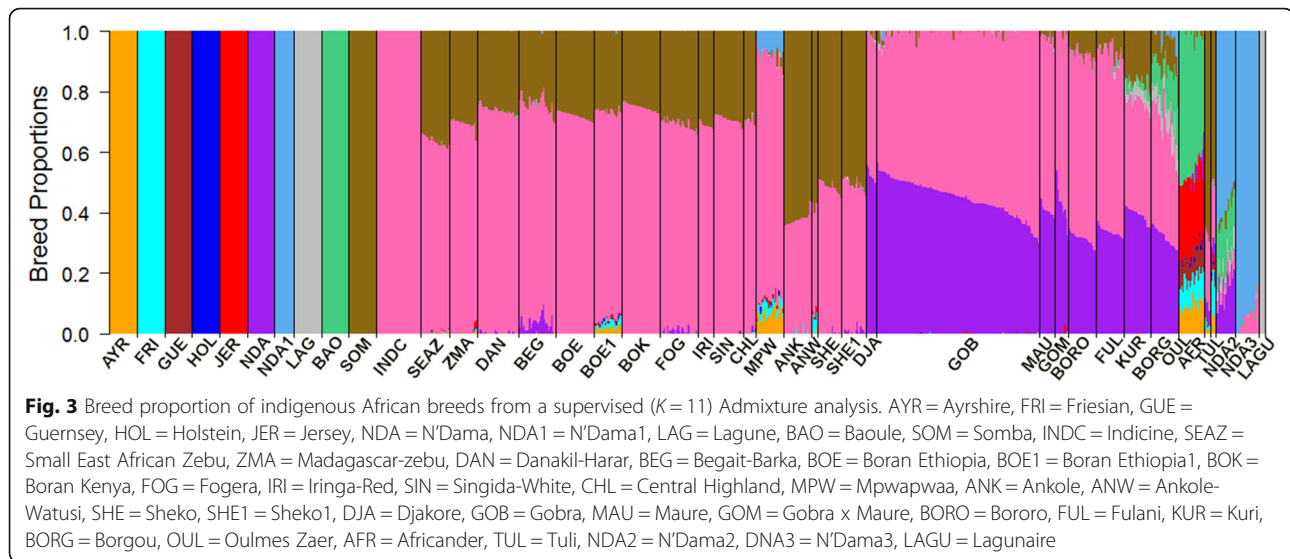
African Tuli and Africander had low indicine ancestry with high levels of exotic taurine ancestry (32 and 10%, respectively; Fig. 1, Table 1).

The Oulmes Zaer were almost exclusively of taurine ancestry with 38.8% African and 60% European taurine ancestry. The synthetic Mpwapwa breed had European taurine (12%) and indicine (87%) ancestry (Fig. 2, Table 1). The African taurine breeds N'Dama1, Lagune, Lagunaire, and Baoule showed >99% reference N'Dama ancestry, whereas N'Dama2 and N'Dama3 (sampled from Southeast and Southwest Burkina Faso, respectively), and Somba showed a high indicine ancestry (8.4, 11.8 and 4.4% indicine, respectively) (Fig. 2, Table 1).

Admixture Model 2, which included five African *Bos taurus* breeds as ancestral reference breeds, identified a difference in the assigned African taurine ancestry between cattle breeds from East, South, and West Africa. The East and South African breeds had a Somba background predominantly. Begait-Barka was the only East African breed with more than 1% N'Dama ancestry. The two South African breeds, Africander and Tuli, showed 6% N'Dama ancestry, while the West African indigenous breeds had N'Dama background (Fig. 3, Table 1) predominantly. However, Bororo, Fulani, Kuri, and Borgou also showed some Somba ancestry (7.8, 6.2, 16.7, and 9.9%, respectively), and the latter two also showed a Lagune background of 3.5 and 6.1%, respectively (Fig. 3, Table 1). Borgou showed an additional N'Dama1 content of 2.4%.

Under Model 2, the African taurine proportion of Oulmes Zaer was 45% Lagune and 2% N'Dama ancestry. The European breed proportion of Mpwapwa remained almost unchanged, but the indicine content was reduced, and African taurine content of 7% N'Dama1 and 1% Somba was detected (Fig. 3, Table 1). Of the African





taurine breeds that were not used as reference ancestral breeds, N'Dama2 appeared to be an admixture of all reference African taurine breeds (N'Dama1 = 58.8%, Baoule = 15.7%, N'Dama = 15.3%, Somba = 2.1%, and Lagune = 4.0%) plus 4.1% *Bos indicus* ancestry. N'Dama3 showed 94.2% N'Dama1 plus 5.2% *Bos indicus* ancestry. The Lagunaire appeared as 100% Lagune (Fig. 3, Table 1).

Admixture and principal components analyses of crossbred cattle

Principal components and Admixture analyses were conducted, including East (Kenya, Uganda, Ethiopia, and Tanzania) and West (Senegal) African crossbred animals. Admixture Model 3 with $K = 12$ extended Model 2 by adding Montbeliarde as a reference breed due to its reported use in crossbreeding in Senegal [21].

Figure 4 shows the PC plots for the same analyses as in Fig. 1 but with crossbred animals added to the plot. The crossbreds from Ethiopia, Kenya, and Tanzania were distributed between the East African zebu and European dairy breeds, while the crossbred animals from Uganda were located between the Ugandan Sanga breed (Ankole) and the European dairy breeds (Fig. 4a). The Senegal crossbred animals exhibited a much greater genetic diversity with a much wider range of both indigenous and exotic dairy breed ancestries compared to the East African crossbreds (Fig. 4b).

With Admixture Model 3, the crossbred animals from Kenya showed an average exotic dairy proportion of 69%, mainly from Friesian (23%), followed by Ayrshire (23%), and Guernsey (16%). The Ugandan crossbreds showed an average exotic dairy proportion of 62% with the main contribution from

Friesian (38%) and Holstein (14%). The Ethiopian and Tanzanian crossbred animals showed an average of 72% exotic dairy proportion; Ethiopian crossbreds had 36% Friesian and 30% Holstein, and Tanzanian crossbreds had 31% Friesian, 19% Ayrshire, and 12% Holstein (Fig. 5, Table 2). The Senegal crossbreds had an average exotic dairy proportion of 50%, ranging from almost 0 to 98%, and the primary contributions coming from Montbeliarde (22%) and Holstein (12%) (Fig. 5, Table 2).

Genetic relatedness and differentiation

Inbreeding, as represented by the F_{IS} value, was close to zero (between -0.006 to 0.009) for most breeds across all breed groups (Table S1). The highest positive F_{IS} value of 0.049 was observed for Somba. The strongest negative F_{IS} was observed for N'Dama3 (-0.109).

Breed differentiation, as represented by F_{ST} values, showed a strong divergence within different groups of breeds (European *Bos taurus*, African *Bos taurus*, zebu types, Sanga types including admixed breeds, and *Bos indicus*; Fig. 6, Table S1). Ranked from lowest to highest genetic differentiation between breeds within groups are zebu types, *Bos indicus*, Sanga types, African *Bos taurus*, and lastly, European *Bos taurus*. Some notable outliers within the breed groups are N'Dama3, which has high F_{ST} with all other Africa *Bos taurus* breeds; Madagascar-zebu with high F_{ST} values with all other zebu type breeds; the South African Africander and Tuli both have high F_{ST} with Sanga type breeds; and Ankole-Watusi which has a relatively high F_{ST} with other Sanga breeds.

Extent and decay of linkage disequilibrium

The decay of squared correlations (r^2) and adjusted squared correlation (r^2_{adj}) between phased alleles of

Table 1 Admixture proportions from supervised analyses (mean \pm SD) of African indigenous breeds for indicine, African taurine and total European taurine ancestry

Breed	K = 7			K = 11				
	Indicine	N'Dama	EUT	Indicine	N'Dama	Somba	EUT	NDA1 + LAG+BAO
Baoule	0.01 \pm 0.02	0.99 \pm 0.02	0	Fixed ancestral breeds				
Somba	0.04 \pm 0.01	0.95 \pm 0.01	0					
Lagune	0	0.99 \pm 0.00	0					
N'Dama1	0	0.99 \pm 0.00	0					
N'Dama2	0.08 \pm 0.03	0.93 \pm 0.03	0	0.04 \pm 0.02	0.15 \pm 0.05	0.02 \pm 0.02	0	0.79 \pm 0.12
N'Dama3	0.12 \pm 0.04	0.88 \pm 0.04	0	0.05 \pm 0.04	0	0	0	0.95 \pm 0.06
Lagunaire	0	0.99 \pm 0.00	0	0	0	0	0	0.99 \pm 0.00
Africander	0.42 \pm 0.01	0.48 \pm 0.02	0.10 \pm 0.02	0.23 \pm 0.01	0.06 \pm 0.01	0.67 \pm 0.02	0.04 \pm 0.03	0
Tuli	0.32 \pm 0.02	0.36 \pm 0.02	0.32 \pm 0.04	0.18 \pm 0.02	0.06 \pm 0.02	0.49 \pm 0.01	0.26 \pm 0.07	0.01 \pm 0.01
Djakore	0.57 \pm 0.02	0.43 \pm 0.02	0	0.46 \pm 0.02	0.52 \pm 0.02	0.01 \pm 0.01	0	0
Gobra	0.65 \pm 0.06	0.35 \pm 0.06	0	0.55 \pm 0.06	0.45 \pm 0.06	0.00 \pm 0.01	0	0
Maure	0.68 \pm 0.02	0.32 \pm 0.02	0	0.57 \pm 0.02	0.41 \pm 0.02	0.02 \pm 0.02	0	0
GobraxMaure	0.66 \pm 0.06	0.33 \pm 0.07	0.01 \pm 0.01	0.56 \pm 0.06	0.43 \pm 0.07	0.01 \pm 0.01	0.01 \pm 0.01	0
Bororo	0.73 \pm 0.02	0.27 \pm 0.02	0	0.61 \pm 0.02	0.31 \pm 0.02	0.08 \pm 0.02	0	0
Fulani	0.71 \pm 0.03	0.29 \pm 0.03	0	0.60 \pm 0.03	0.34 \pm 0.02	0.06 \pm 0.02	0	0.01 \pm 0.01
Kuri	0.50 \pm 0.02	0.50 \pm 0.02	0	0.38 \pm 0.02	0.40 \pm 0.02	0.17 \pm 0.02	0	0
Borgou	0.47 \pm 0.07	0.53 \pm 0.07	0	0.37 \pm 0.06	0.32 \pm 0.03	0.10 \pm 0.02	0	0.21 \pm 0.01
Ankole	0.61 \pm 0.02	0.36 \pm 0.02	0.03 \pm 0.02	0.37 \pm 0.02	0	0.62 \pm 0.02	0.00 \pm 0.01	0
Ankole-Watusi	0.55 \pm 0.02	0.32 \pm 0.02	0.13 \pm 0.04	0.35 \pm 0.02	0	0.59 \pm 0.02	0.06 \pm 0.05	0
SEAZ	0.81 \pm 0.02	0.18 \pm 0.01	0.01 \pm 0.01	0.64 \pm 0.02	0	0.36 \pm 0.02	0	0
Madagascar-zebu	0.85 \pm 0.03	0.15 \pm 0.01	0.01 \pm 0.02	0.69 \pm 0.03	0	0.31 \pm 0.02	0.01 \pm 0.02	0
Danakil-Harar	0.88 \pm 0.01	0.12 \pm 0.01	0	0.74 \pm 0.01	0	0.26 \pm 0.01	0	0
Begait-Barka	0.85 \pm 0.04	0.14 \pm 0.04	0.01 \pm 0.04	0.72 \pm 0.03	0.03 \pm 0.02	0.24 \pm 0.03	0.01 \pm 0.01	0.00 \pm 0.01
Boran Ethiopia	0.87 \pm 0.01	0.13 \pm 0.01	0	0.72 \pm 0.01	0	0.28 \pm 0.01	0	0
Boran Ethiopia1	0.83 \pm 0.02	0.12 \pm 0.01	0.06 \pm 0.01	0.69 \pm 0.01	0	0.26 \pm 0.01	0.05 \pm 0.01	0
Boran Kenya	0.90 \pm 0.01	0.10 \pm 0.01	0	0.75 \pm 0.01	0	0.25 \pm 0.01	0	0
Fogera	0.84 \pm 0.01	0.16 \pm 0.01	0	0.69 \pm 0.01	0	0.30 \pm 0.02	0.00 \pm 0.02	0
Iringa-Red	0.85 \pm 0.03	0.14 \pm 0.01	0.01 \pm 0.03	0.68 \pm 0.02	0	0.31 \pm 0.01	0.01 \pm 0.02	0
Singida-White	0.87 \pm 0.01	0.13 \pm 0.01	0	0.71 \pm 0.02	0	0.29 \pm 0.01	0	0
Central Highland	0.85 \pm 0.13	0.15 \pm 0.01	0.00 \pm 0.01	0.70 \pm 0.02	0	0.30 \pm 0.012	0	0
Sheko	0.67 \pm 0.02	0.32 \pm 0.01	0.01 \pm 0.01	0.48 \pm 0.02	0.00 \pm 0.01	0.51 \pm 0.02	0	0
Sheko1	0.67 \pm 0.02	0.32 \pm 0.01	0.01 \pm 0.01	0.49 \pm 0.02	0.01 \pm 0.01	0.50 \pm 0.02	0	0
Mpwapwa	0.87 \pm 0.03	0.01 \pm 0.02	0.12 \pm 0.03	0.81 \pm 0.03	0	0.01 \pm 0.02	0.11 \pm 0.05	0.07 \pm 0.02
Oulmes Zaer	0.01 \pm 0.01	0.38 \pm 0.04	0.60 \pm 0.04	0	0.02 \pm 0.03	0	0.52 \pm 0.09	0.45 \pm 0.05

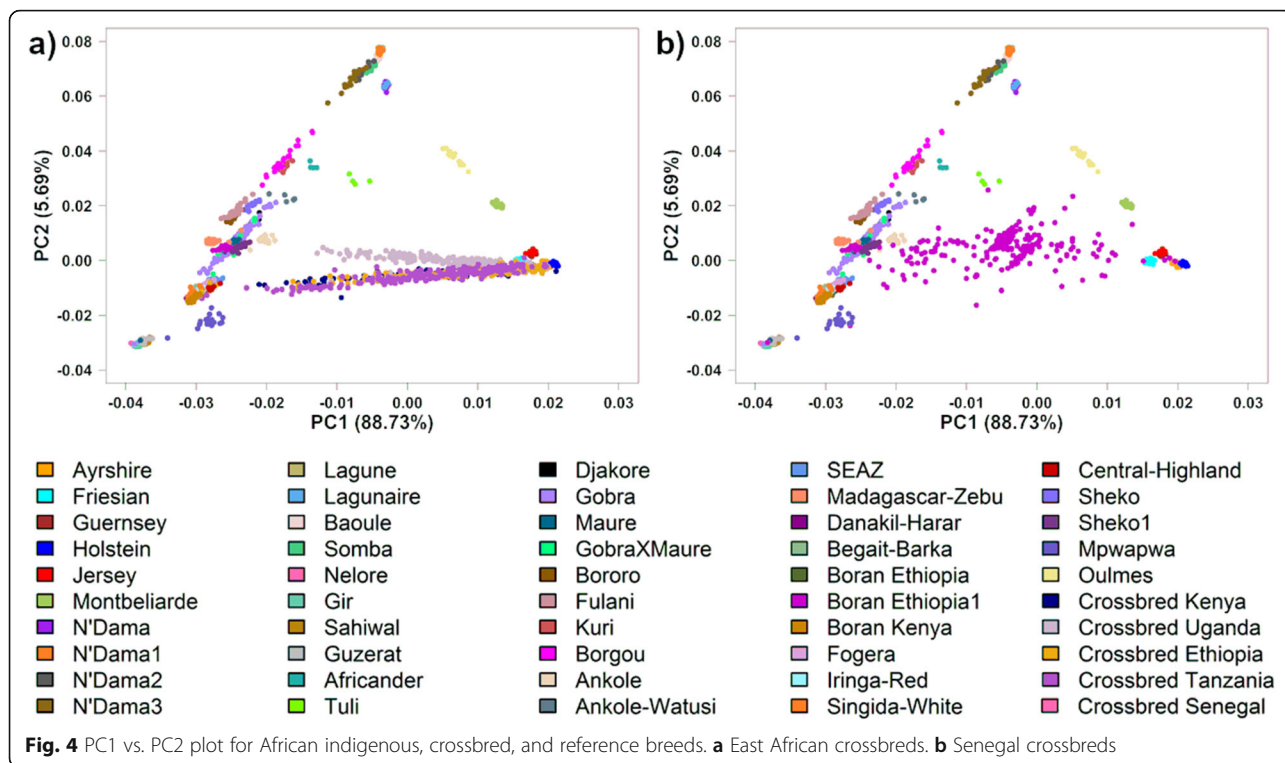
EUT 5 European *Bos taurus* breeds, NDA1 N'Dama1, LAG Lagune, BAO Baoule

pairwise SNP loci over increasing genome distances is illustrated in Fig. 7a and b, respectively, for the nine African indigenous breeds that had more than 20 animals after removing highly related animals from the data. Ankole had higher r^2 and a lower rate of r^2 decay, and Gobra showed lower r^2 and a higher rate of r^2 decay than the other populations across all distances (Fig. 7a),

which translates into the lowest and highest estimates of N_e across all times, respectively.

Past effective population size before and after adjusting r^2 for sample size

N_e was calculated for various generations in the past using r^2 and r^2_{adj} for the nine African indigenous breeds



in the analyses. N_e estimates using r^2 declined steadily over time for all breeds (Fig. 8a). Except for Ankole, which showed a steady decline across all periods, the N_e estimates using r^2_{adj} declined until around 200 generations ago and then held steady or increased markedly until 30 to 5 generations ago before declining again.

Gobra showed the highest and Ankole the lowest N_e at all generations using r^2 , with 107 and 18 at 1 generation ago, and 6544 and 4633 at 2000 generations ago, respectively. Similarly, estimates of N_e based on r^2_{adj} for Gobra were highest at generation 1 and 2000, with 3418 and 6809, respectively, while the lowest N_e was found for Ankole, with 272 at generation 1 and 5557 at generation 2000. Estimated N_e using r^2 for Bororo, SEAZ, Danakil-Harar, Fogera, Boran Ethiopia, and Begait-Barka

at 1 and 2000 generations ago were 19 and 4687, 19 and 4812, 20 and 5005, 24 and 5063, 25 and 5168, 25 and 5255, respectively, while estimates of N_e using r^2_{adj} were 743 and 5630, 576 and 5790, 410 and 6006, 665 and 5899, 363 and 5964, and 659 and 6109, respectively. Thus, across the nine breeds, the finite sampling adjustment to r^2 increased N_e 14.8 to 39.2 fold at generation 1 and 1 to 1.2 fold at 2000 generations ago.

Discussion

Genetic diversity and relationships

Depending on the used data and underlying assumption about biological clocks, estimates of divergence between *Bos taurus* and *Bos indicus* vary from approximately 200,000 to 300,000 years BP [3, 14, 16, 22, 23], to 575,000 to

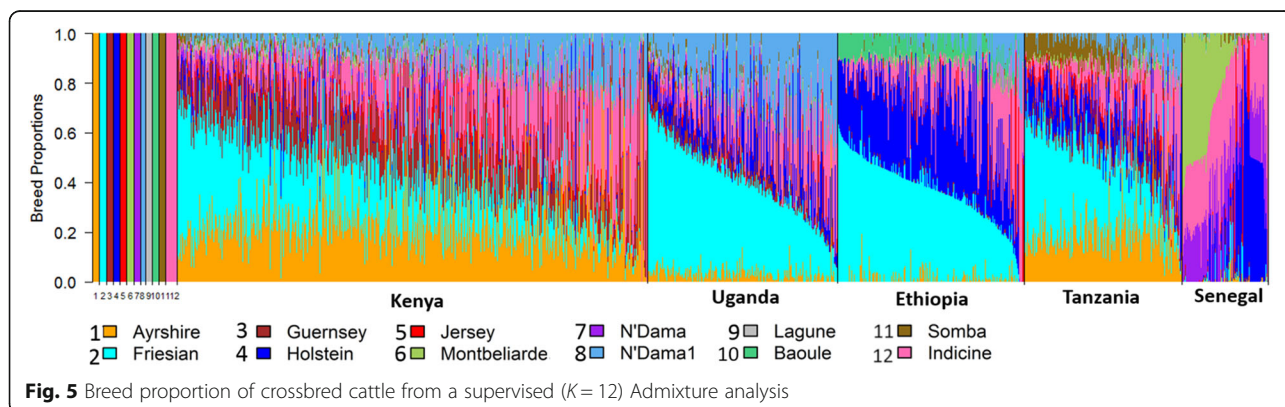


Table 2 Estimates (\pm SD) of individual and total exotic dairy breed proportions for African crossbred cows

Breed	Total dairy proportion	Ayrshire	Friesian	Guernsey	Holstein	Jersey	Montbeliarde
Kenyan Crossbred	0.688 \pm 0.202	0.228 \pm 0.127	0.230 \pm 0.126	0.157 \pm 0.084	0.030 \pm 0.091	0.036 \pm 0.043	0.008 \pm 0.009
Ugandan Crossbred	0.622 \pm 0.195	0.033 \pm 0.041	0.381 \pm 0.141	0.048 \pm 0.052	0.137 \pm 0.132	0.015 \pm 0.032	0.007 \pm 0.009
Ethiopian Crossbred	0.721 \pm 0.190	0.014 \pm 0.024	0.360 \pm 0.133	0.004 \pm 0.008	0.300 \pm 0.139	0.036 \pm 0.109	0.006 \pm 0.009
Tanzanian Crossbred	0.719 \pm 0.172	0.188 \pm 0.092	0.307 \pm 0.122	0.046 \pm 0.042	0.121 \pm 0.095	0.046 \pm 0.032	0.011 \pm 0.014
Senegalese Crossbred	0.503 \pm 0.187	0.006 \pm 0.016	0.016 \pm 0.037	0.008 \pm 0.018	0.222 \pm 0.255	0.026 \pm 0.103	0.224 \pm 0.227

800,000 years BP [15, 24], to 2 m years BP [25]. Based on microsatellite data, MacHugh et al. [19] estimated the separation of the African and European taurine clades to be between 180,000 to 250,000 years ago, while the same group estimated divergence time at between 22,000 and 26,000 years ago using mtDNA [15]. Both estimates pre-date domestication.

The PCA and F_{ST} results showed a clear divergence between *Bos indicus*, African *Bos taurus*, and European *Bos taurus* reference breeds, which is in agreement with several previous studies [4, 26–32]. While the pattern of PCA results is expected to reflect in part the choice of

SNPs on the assay, we have found very similar patterns for those breeds that have Illumina 777 k SNP data, despite the widely different pattern of breed allele frequency distributions of the SNPs on the 777 k versus the 50 k assays [33]. As shown here and in other studies, the first two principal components differentiate the groups with the largest genetic differences (PC1 *Bos taurus* vs. *Bos indicus*, PC2 European *Bos taurus* vs. Africa *Bos taurus*). Later principale components could be able to tease out smaller genetic differences, such as dairy vs beef breeds, however, coresponding beef reference breeds were not included. Additionally, some beef breeds are closer

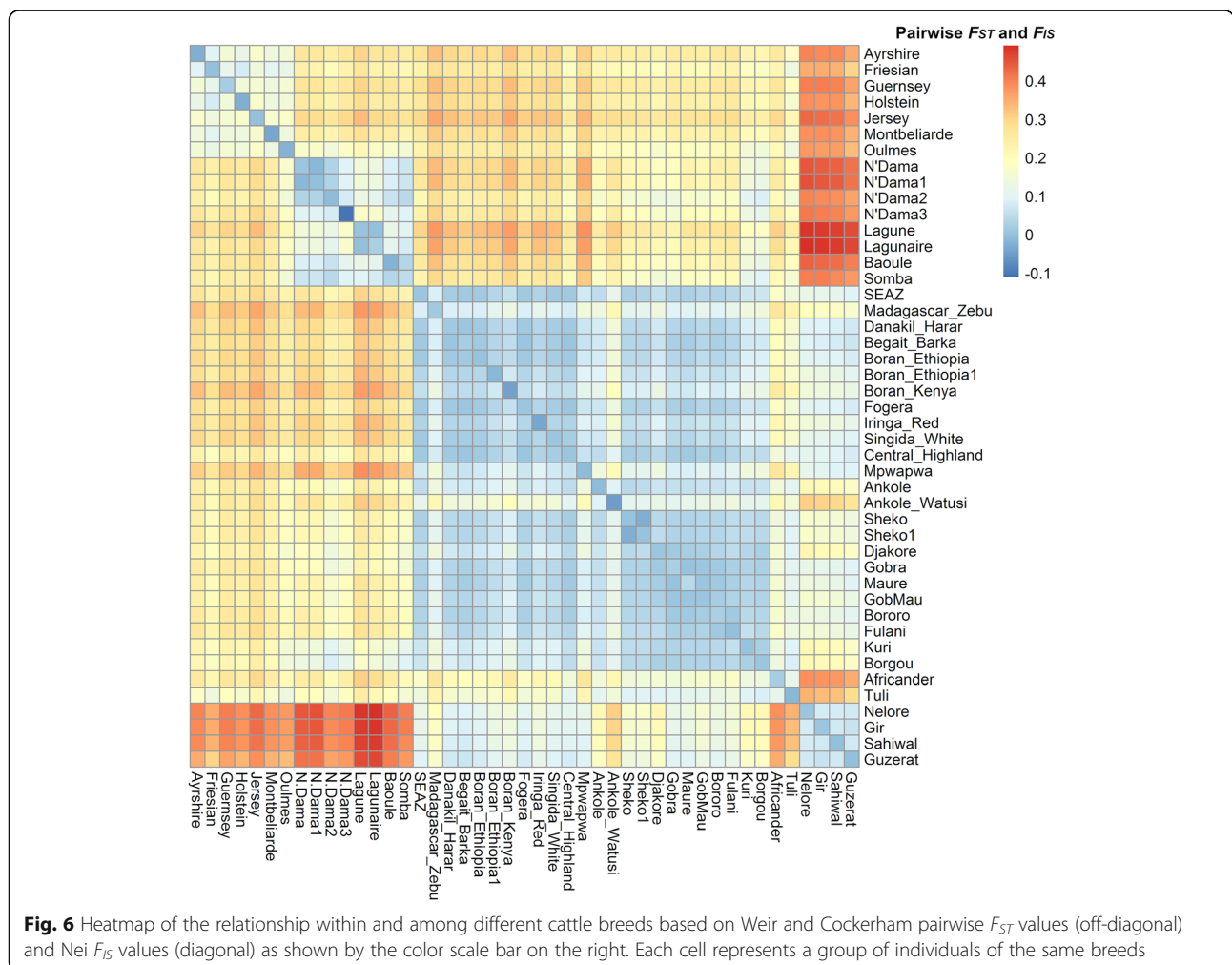
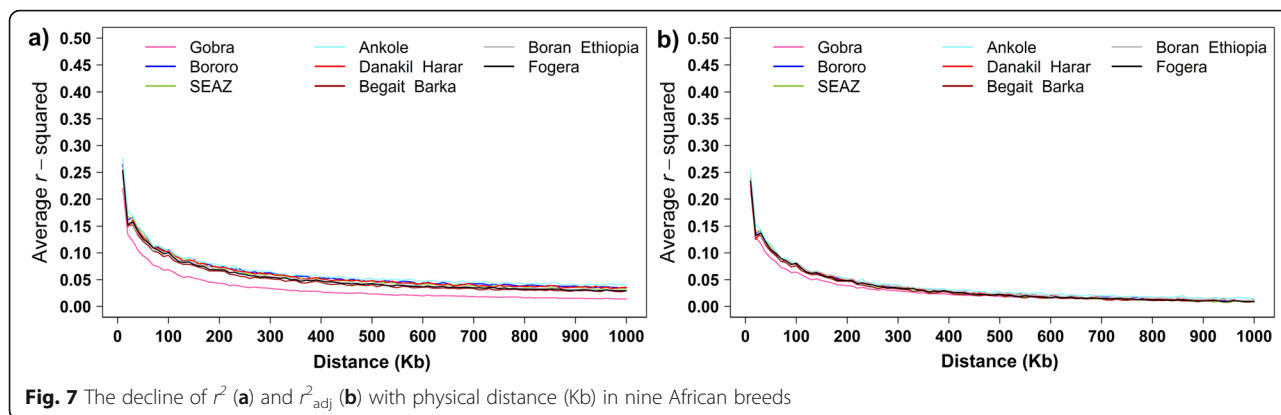


Fig. 6 Heatmap of the relationship within and among different cattle breeds based on Weir and Cockerham pairwise F_{ST} values (off-diagonal) and Nei F_{IS} values (diagonal) as shown by the color scale bar on the right. Each cell represents a group of individuals of the same breeds



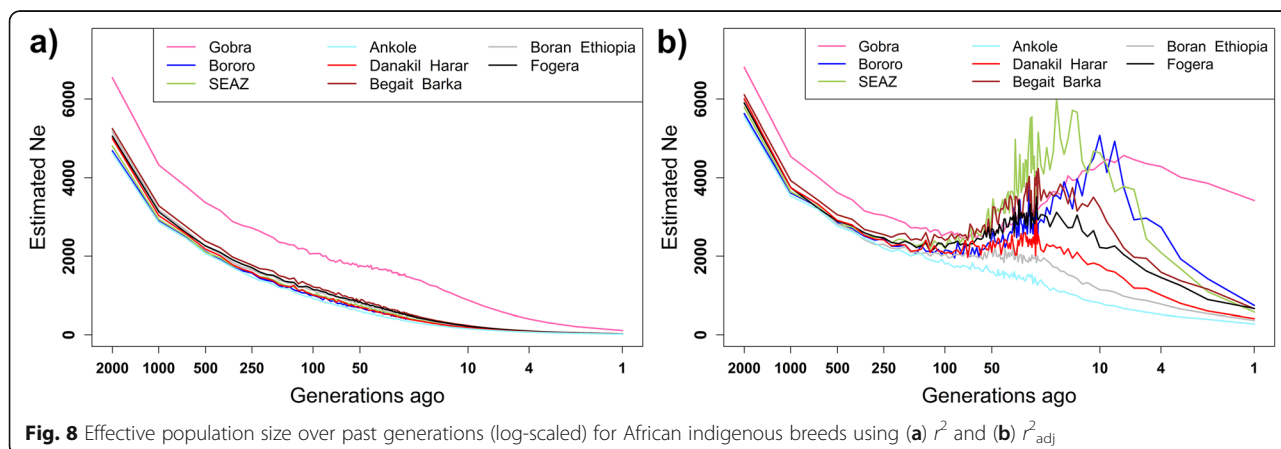
related to dairy breeds than other beef breeds or vice versa [34–36] which can make the choice of appropriate reference breeds cumbersome. As we were interested in the origin of African cattle and not their economic type, we chose to only include reference breeds that are historically proven to have been introduced in the regions under study.

Among the African cattle populations, there are no pure *Bos indicus* populations in our sample. Based on the range of samples included in our study, it is unlikely that any indigenous breeds of Africa are pure *Bos indicus*. As outlined in the background section, it has been assumed in the literature that the admixed indigenous breeds of Africa arose from *Bos indicus* cattle entering Africa and breeding with existing African *Bos taurus* populations [8, 9]. Verdugo et al. [4] concluded that *Bos taurus* populations in the Near East became admixed with *Bos indicus* likely due to human migrations around 4200 years BP. The populations they sampled in the Levant, which were the closest samples to the putative first route of *Bos indicus* into Africa through Egypt, showed a *Bos taurus* genotype that was closest to modern African *Bos taurus*. Therefore, it is possible that the first humped cattle entering Africa around 4000 years BP [6] may

already have been hybrids between *Bos taurus* and *Bos indicus* rather than pure *Bos indicus*.

Payne and Hodges [6] state that it is likely that many African cattle were *Bos taurus* before the rinderpest virus (RPV) epidemic of 1887–1897, which was reported to have more severely affected the taurine cattle populations of the East and South than the zebu populations. However, inferences about the types of cattle present in Africa until recent times are largely based on sparse depictions of humped (inferred as *Bos indicus* types) versus non-humped (inferred as *Bos taurus* type) cattle.

Our results show that, other than the pure African *Bos taurus* populations, all African indigenous cattle populations are admixtures of *Bos indicus* and African *Bos taurus*, with West African and Southern African populations showing lower *Bos indicus* admixture than East African populations consistent with recent studies by Pitt et al. [37] and Verdugo et al. [4]. The South African breeds, Africander and Tuli, show an even higher differentiation from East and West African breeds, as well as from other Sanga breeds based on F_{ST} values. This differentiation can be attributed at least in part to their admixture with European taurine breeds. Some indigenous breeds such as the Gobra show a surprisingly wide distribution of African taurine and indicine ancestry (Fig.



1a, Ndiaye et al. [38]), which likely reflects recent crossing to Guzerat (a pure *Bos indicus* breed) that was imported from Brazil into the region in which the Gobra samples used here were collected [39].

The PCA showed that the existing African *Bos taurus* populations and our reference European *Bos taurus* populations exhibit much greater diversity than the set of *Bos indicus* populations used in our analyses. Studies that included much larger samples of European *Bos taurus* cattle similarly showed greater diversity among European *Bos taurus* versus *Bos indicus* cattle (e.g. Decker et al., [40] and Mastrangelo et al., [41]). Our results indicate that the greater diversity of *Bos taurus* cattle extends to African *Bos taurus* in addition to European *Bos taurus*.

Populations with relatively high heterozygosity levels were Ankole-Watusi and N'Dama3, which show substantial admixture between populations with a large genetic distance (indigenous with European *Bos taurus*, and African *Bos taurus* with other indigenous, respectively). The distribution of indigenous breeds in the PC plots (Fig. 1a, b) and the results from Admixture (Table 1, $K=11$) suggest a different African *Bos taurus* ancestry for West African versus East African indigenous.

The West African Bororo (sampled in Chad) clustered closely with the Fulani (sampled in Benin) and had very similar breed proportions, indicating that they likely form a single population. According to Grema et al. [42], Bororo is also known by several other names, including Red Fulani in West Africa supporting that Bororo and Fulani are one population with different names in different countries. Kuri is generally referred to as being an African *Bos taurus* breed [42], but our results show that Kuri is an admixture between African *Bos taurus* and *Bos indicus*. The tight clustering of the Kuri samples indicates that this is an old admixture rather than a recent hybridization.

Results for the Mpwapwa showed that this synthetic dual-purpose breed, first created about 60 years ago, aligns with its reported genetic history of 35% Sahiwal, 20% Red Sindhi, 35% East African zebu breeds, and 10% Ayrshire [43]. The Oulmes Zaer showed high heterozygosity consistent with its high European *Bos taurus* ancestry and being an admixture between African and European *Bos taurus*. Gautier and Naves [44], who analyzed the same Oulmes Zaer samples as here but with a very different set of 23 other breeds, also found that Oulmes Zaer was of hybrid origin between African and European *Bos taurus*. The rather dispersed cluster of the Oulmes Zaer samples in the PC plots suggests the possibility that the breed might have been deliberately created by the crossing of African and European *Bos taurus* cattle in relatively recent times.

Crossbreeding is widely used to achieve a suitable balance of productivity and adaptation in African smallholder dairy farming. According to the literature, the predominant exotic dairy breeds used in Kenya have been Holstein-Friesian, Ayrshire, Guernsey, and Jersey [45], and in Uganda Holstein-Friesian, with some use of Jersey, Guernsey, and Ayrshire [46]. In Ethiopia and Tanzania, the dominant exotic breeds are Holstein-Friesian and Jersey, and Holstein and Friesian, respectively [47, 48], while in Senegal Montbeliarde and Holstein are the dominant exotic breeds [49]. In the current study, the European dairy proportions found in East and West African crossbreds reflect the reported history of crossbreeding. Based on the PC plot (Fig. 4), the crossbred animals from Kenya, Ethiopia, and Tanzania distributed towards East African zebu, the dominant indigenous breed group in these countries, while in Uganda, they distributed toward the Ankole, which was the dominant indigenous breed in the areas sampled. These results are consistent with the findings of Weerasinghe et al. [13] and Strucken et al. [33] for the East African crossbred samples used here.

Linkage disequilibrium and effective population size

Linkage disequilibrium is a measure of the non-random association of alleles at two or more loci that can be caused by selection and past and present population structure [50, 51]. As expected, the LD for all populations in our study was the highest ($r^2_{\text{adj}} > 0.2$) for very short distances (about 10 Kb) with an exponential decline with increasing distance. The European reference dairy breeds showed a higher level of LD across all interval sizes compared to *Bos indicus* reference breeds and most African indigenous breeds (Figure S3), reflecting that intensive selection has caused relatively low effective population sizes generating high LD in the European dairy breeds [52, 53].

We found for all populations a steady, slow decline in N_e from 2000 generations until about 150 generations ago. Assuming a generation interval of about 6 years, this time range corresponds to about 12,000–900 years ago. The domestication of cattle has been dated to about 10,000 years ago [54, 55]. The *Bos indicus* reference breeds and all African indigenous populations, including the African *Bos taurus* breeds, show a substantially higher N_e 12,000 years ago ($N_e = 5000$ to 7000) than the European *Bos taurus* breeds ($N_e = 2000$ to 3500, Figure S4). Given the admixed *Bos indicus* and African *Bos taurus* ancestry of most African indigenous breeds, a high N_e pre-domestication is not unexpected, but the higher N_e of *Bos indicus* and African *Bos taurus* compared to European *Bos taurus* suggest that European *Bos taurus* was domesticated from a smaller population than African *Bos taurus* and *Bos indicus*. The N_e estimates

obtained using r^2_{adj} for recent generations for all indigenous populations are > 100 , which should provide sufficient genetic diversity for the long-term survival of a population according to Meuwissen [56].

Estimates of N_e based on r^2_{adj} stabilized or increased for most African indigenous breeds between about 200 generations and 30 to 10 generations ago (Figs. 8b, S5c and d). The process of domestication and stabilization of populations into genetically discrete populations could lead to lower effective population sizes over time, even if numbers of cattle overall were increasing with the spread of cattle farming. But as human populations and their associated livestock populations continued to expand, growth in numbers of cattle within each indigenous population could have outweighed downward pressure on N_e . An alternative explanation of the increase in N_e about 200 generations ago might be a period of hybridization between existing indigenous populations and/or between indigenous populations and *Bos indicus* cattle, following the migration of Bantu-speaking agropastoralists from eastern to southern Africa around 1500 years BP [57], and the second wave of *Bos indicus* cattle that are believed to have entered Africa with the migration of Arab peoples starting around 1500 to 1300 years BP.

Finally, a stabilization or increase in N_e might arise if the adjustment of r^2 to account for sample size led to over-correction of r^2 causing underestimation of r^2_{adj} and hence overestimation of N_e . But in this case, the bias in N_e should be more pronounced for the most recent generations where N_e is estimated from very long-range estimates of LD, which have the lowest expectations of true r^2 and hence r^2_{adj} values that are closest to zero leading to high values of N_e . The expectation from theory is that very large sample sizes are required to avoid downward bias in the estimation of N_e [58]. Corbin et al. [59] confirmed this using simulation to show that before adjusting r^2 values for sampling effect, increasing the sample size reduced the downward bias and improved the accuracy of estimates of N_e for recent generations. However, when r^2 values were adjusted for sample size, the estimates of N_e were more stable. These authors suggested that the theoretical basis for models of variable N_e based on LD is not clear and has not been fully established. Therefore, the conservation decisions based on N_e from LD should be considered with caution.

Conclusions

African indigenous cattle are genetically diverse due to historical and highly diverse admixture of *Bos indicus*, African *Bos taurus*, and European *Bos taurus*. This provides a great opportunity for future research and utilization, particularly for traits underlying adaptation

to challenging environments. Our study shows that historical classifications of breeds only approximate the underlying genetic differences, with substantial overlap in the composition of breed groups historically thought to be distinct, such as Sanga and zebu. We also showed that African zebu cattle are not *Bos indicus* as continues to be stated in much of the published literature. The African continent is home to many more breeds than analyzed here, and it is to be hoped that more extensive application of molecular genetic assays will lead to the characterization of all the breeds of Africa yielding a comprehensive map of African cattle diversity, creating a platform for future utilization and characterization.

Methods

Animals and their sources

This study analyzed data on 4089 animals representing a diverse set of African indigenous and crossbred plus exotic cattle populations. In addition to the crossbreds, exotic reference breeds, one African synthetic breed, and one unclassified African indigenous breed were used, and 669 samples representing 33 breeds or populations within the three main African indigenous cattle breed groups of African *Bos taurus*, African zebu, and Sanga (Table 3). Data were obtained from several public-domain databases plus two projects run by the International Livestock Research Institute (ILRI) and collaborators, and the Dairy Genetics East Africa project (DGEA, Strucken et al. [33]). The main objective of DGEA was to identify the most suitable crossbred dairy cow genotypes for the range of dairy production systems and levels of production in Kenya, Uganda, Tanzania, and Ethiopia. The samples from Senegal [21, 39, 60] were collected as part of a study on the trade-offs of keeping different breed or cross-breed types of dairy cattle in smallholder systems in Senegal. A total of 644 samples were collected, with the breed or cross-breed type of the animal given by the farmer. The main breed types comprised indigenous breeds crosses between the indigenous breeds and the Guzerat (an imported *Bos indicus* dairy breed), and crosses between the indigenous breeds and other exotic dairy breeds, such as Montbeliarde and Holstein-Friesian.

Bos indicus and European *Bos taurus* breeds were included to represent the major known anchor points of global cattle diversity against which African cattle diversity can be assessed. The European *Bos taurus* breeds were chosen because these breeds are known to have contributed to the crossbred dairy populations included in our study, allowing clearer interpretation of results than inclusion of other breeds or a larger sample of breeds. The reference breeds included five African *Bos taurus* (N'Dama, N'Dama1, Lagune, Baoule, and Somba), four *Bos indicus* (Nelore, Sahiwal, Gir, and Guzerat), and

Table 3 Animal populations, numbers, and sources

Breed	Breed group	Geographical location	Origin/Country	Number of animals	Genotype source
Ayrshire	EuB.t	Canada	Canada	20	CDN
Friesian	EuB.t	European	UK	20	SRUC
Guernsey	EuB.t	USA and UK	USA and UK	20	Bovine HapMap consortium [32]
Holstein	EuB.t	USA and NZ	USA and NZ	20	Bovine HapMap consortium [32]
Jersey	EuB.t	USA and NZ	USA and NZ	20	Bovine HapMap consortium [32]
Montbeliarde	EuB.t	European	France	20	Decker et al., [40]
N'Dama	AfB.t	West African	Guinea	20	Bovine HapMap consortium [32]
N'Dama1	AfB.t	West African	Cote d'Ivoire	14	Decker et al. [40]
Lagune	AfB.t	West African	Benin	20	Decker et al. [40]
Somba	AfB.t	West African	Togo	20	Decker et al. [40]
Baoule	AfB.t	West African	Burkina Faso	20	Decker et al. [40]
Nelore	B.i	Brazil	Brazil	20	Bovine HapMap consortium [32]
Gir	B.i	Indian	Brazil	20	DGEA [33]
Sahiwal	B.i	Indian	Kenya	20	DGEA [33]
Guzerat*	B.i	Brazil	Senegal	8	Marshall et al. [39]
N'Dama2	AfB.t	West African	Southeast Burkina Faso	14	Decker et al. [40]
N'Dama3	AfB.t	West African	Southwest Burkina Faso	17	Decker et al. [40]
Lagunaire	AfB.t	West African	West Africa	5	Bovine HapMap consortium [32]
Africander	Sanga	Southern African	South Africa	4	Decker et al. [40]
Tuli	Sanga	Southern African	Botswana	4	Decker et al. [40]
Djakore	zebu	West African	Senegal	7	Marshall et al. [39]
Gobra	zebu	West African	Senegal	118	Marshall et al. [39]
Maure	zebu	West African	Senegal	12	Marshall et al. [39]
Gobara*Maure	zebu	West African	Senegal	10	Marshall et al. [39]
Bororo	zebu	West African	Chad	20	Decker et al. [40]
Fulani	zebu	West African	Benin	20	Decker et al. [40]
Kuri	AfB.t	West African	Chad	20	Decker et al. [40]
Borgou	zebu	West African	Benin	20	Decker et al. [40]
Ankole	Sanga	East African	Uganda	20	DGEA [33]
Ankole-Watusi	Sanga	East African	Ruanda	5	Decker et al. [40]
SEAZ	zebu	East African	Kenya	21	DGEA [33]
Madagascar-zebu	zebu	Madagascar	Madagascar	20	Decker et al. [40]
Danakil-Harar	zebu	East African	Ethiopia	30	DGEA [33]
Begait-Barka	zebu	East African	Ethiopia	27	DGEA [33]
Boran	zebu	East African	Ethiopia	28	DGEA [33]
Boran1	zebu	East African	Ethiopia	20	Decker et al. [40]
Boran	zebu	East African	Kenya	28	DGEA [33]
Fogera	zebu	East African	Ethiopia	28	DGEA [33]
Iringa-Red	zebu	East African	Tanzania	11	DGEA [33]
Singida-White	zebu	East African	Tanzania	22	DGEA [33]
Central Highland	zebu	East African	Ethiopia	9	DGEA [33]
Sheko	Sanga	East African	Ethiopia	17	Decker et al. [40]
Sheko1	Sanga	East African	Ethiopia	18	Bovine HapMap consortium [32]
Mpwapwa	Synthetic	East African	Tanzania	20	DGEA [33]

Table 3 Animal populations, numbers, and sources (Continued)

Breed	Breed group	Geographical location	Origin/Country	Number of animals	Genotype source
Oulmes Zaer	Indigenous	North African	Morocco	19	Decker et al. [40]
Kenyan crossbred	Crossbred	East African	Kenya	1378	DGEA [33]
Uganda crossbred	Crossbred	East African	Uganda	555	DGEA [33]
Ethiopia crossbred	Crossbred	East African	Ethiopia	545	DGEA [33]
Tanzania crossbred	Crossbred	East African	Tanzania	462	DGEA [33]
Senegal crossbred	Crossbred	West African	Senegal	253	Marshall et al. [39]
Total				4089	

*Guzerat was imported from Brazil; *EuB.t* European *Bos taurus*, *AFB.t* African *Bos taurus*, *Bi* *Bos indicus*, *SEAZ* Small East African Zebu, *USA* United States of America, *UK* United Kingdom, *NZ* New Zealand, *SRUC* Scottish Rural University College, *CDN* Canadian Dairy Network; Crossbreds are crosses between *EuB.t.* dairy breeds and local indigenous breeds; synthetic is a breed created by the deliberate crossing of several founder breeds; Indigenous here refers to a single African breed that does not fit into the other classifications

six European *Bos taurus* dairy breeds (Guernsey, Holstein, Jersey, Canadian Ayrshire, British Friesian, and Montbeliarde). The reference data were obtained from the Bovine HapMap Consortium [32], Decker et al. [40], the Scottish Rural University College (SRUC), and the Canadian Dairy Network (CDN, Table 3).

Genotyping and quality control

The Senegal animals were genotyped using the Illumina BovineSNP50v2 BeadChip array (Illumina Inc., San Diego, CA, USA) comprising 54,609 SNPs. Quality control was carried out using the GenABEL package [61] in R Core Team [62]. Autosomal SNPs were retained, and SNPs and animals with call-rates lower than 90% were excluded. No threshold criteria for minor allele frequency (MAF) or Hardy-Weinberg-Equilibrium (HWE) were applied because low MAF SNPs can provide powerful information for breed differentiation analyses and HWE is not expected to yield reasonable results for some populations due to admixture and small population size. A total of 45,809 SNPs and 628 animals remained after quality control.

Samples from the DGEA project were genotyped with the Illumina BovineHD Beadchip (Illumina Inc., San Diego, CA, USA) and sourced from Strucken et al. [33]. The DGEA data of 777 k SNPs was quality controlled according to similar thresholds as described in West Africa. Approximately 735 k SNPs remained after QC. The 777 k data from the Bovine HapMap Consortium, SRUC, and CDN were also supplied quality controlled. The 50 k data from Decker et al. [40] were genotyped with the Illumina BovineSNP50 BeadChip array (Illumina Inc., San Diego, CA, USA), and supplied post quality control. Merging and retaining only those SNPs present across all datasets resulted in a subset of 38,556 SNPs.

Analysis of the genetic structure

Principal component analyses were performed using a GRM to define the covariance between animals. Two

separate PCA were conducted: 1) where the GRM was based on all cattle populations; 2) the GRM was built using only African taurine breeds. The second GRM was used to evaluate the genetic structure and diversity among the African taurine breeds to select the African *Bos taurus* reference populations to be used in the Admixture analysis. The GRMs were constructed according to the first method of VanRaden [63]. Genotypes were recorded as allele counts of 0, 1, and 2, which were converted to -1, 0, 1 to centre the genotype matrix **M**. Missing genotypes were replaced with the average allele frequencies across all animals for each given SNP. The GRM was then calculated as:

$$GRM = ZZ' / d$$

where the scaling parameter *d* was:

$$d = 2 * \sum (p_i * (1 - p_i))$$

The centred genotype matrix (**Z**) was constructed by subtracting the **P** matrix from the genotype matrix **M**, where **P** = 2 * (*p_i* - 0.5), and *p_i* is the allele frequency at locus *i*.

To investigate the genetic admixture of the indigenous and crossbred cattle populations, a maximum likelihood model implemented in the software ADMIXTURE 1.23 [64] was applied. Supervised analyses were used because unsupervised analyses become unstable as *K* is increased, i.e., individual true ancestry proportions are low, and estimates become uninterpretable and often over-predicted [65]. Also, prior literature [33, 66] and our analyses have previously demonstrated that there are three ancestral populations of African cattle (African *Bos taurus*, European *Bos taurus*, and *Bos indicus*), and the use of supervised analyses allows clear dissection of these contributions. A pooled sample of *Bos indicus* reference (indicine) was created with eight animals per breed because of the four indicine reference breeds clustered very closely together in the PCA (Figure S1a). The African taurine samples showed larger breed differences

compared to the indicine reference breeds (Figure S1b) and were therefore considered as separate reference populations. ADMIXTURE was used in 3 alternative supervised analyses where the number of reference breeds was set to 7, 11, and 12. The assumed ancestral populations in Model 1 ($K = 7$) were: N'Dama (African taurine), Indicine (pooled *Bos indicus* samples), Ayrshire, Friesian, Guernsey, Holstein, and Jersey. In Model 2 ($K = 11$), N'Dama1, Lagune, Baoule, and Somba were added as African taurine reference breeds. In Model 3 ($K = 12$), Montbeliarde was added as another European dairy breed, because Montbeliarde has been used for cross-breeding to indigenous cattle in Senegal.

Genetic relatedness and differentiation

Pairwise F_{ST} values were calculated according to Weir and Cockerham [67], where F_{ST} is defined as the genetic variance between populations expressed as a proportion of the total genetic variance. The degree of inbreeding was inferred from the F_{IS} coefficient calculated according to Nei [68], where the F_{IS} is defined as one minus the observed proportion of heterozygotes divided by the expected proportion of heterozygotes. To explore genetic differentiation among breeds, F_{ST} values were visualized in a heatmap, and the complete-linkage method was used for hierarchical clustering as provided in the R package “pheatmap” [69].

The extent of linkage disequilibrium and effective population size in pure breeds

The extent and magnitude of LD within different breeds were determined using r^2 between phased alleles of pairwise SNP loci. The genotypes were phased using Eagle v2.4 [70], and the LD coefficients were calculated with VCFtools v0.1.15 [71]. The r^2 was estimated separately for each breed between all pairs of SNPs with a distance of up to 50 Mb using markers with minor allele frequency $\geq 5\%$ according to the Hill and Roberson formula [72]:

$$r^2 = \frac{D^2}{f(A) \times f(a) \times f(B) \times f(b)}$$

where,

$$D = f(AB) - f(A)f(B)$$

where $f(AB)$ is the observed frequency of haplotype AB [73], while $f(A)$, $f(a)$, $f(B)$ and $f(b)$ are observed frequencies of alleles A , a , B , and b , respectively. For the purpose of graphical display, the distance of pair-wise LD was binned into 10 Kb intervals up to 1 Mb.

The r^2 values combined with marker distances were used to estimate the approximate N_e at a given time point in the past, assuming a model without mutation, and using the formula of Sved [74]:

$$N_e = \left(\frac{1}{4c}\right) \left(\frac{1}{r^2} - 1\right)$$

where N_e is the effective population size, and c the marker distance in Morgans assuming 1 M = 100Mbp. The time (number of generations ago) at which N_e was estimated as $1/2c$ [52]. For estimating the effective population size, the data were grouped into 80 bins of 25 Kb for SNPs that were up to 2 Mb while, for a distance of more than 2 Mb up to 50 Mb, the number of previous generations (25 to 1) was selected, and the appropriate range of c was calculated. The binning process was designed to ensure sufficient SNP pairs within each bin and to obtain a representative average r^2 across the autosomes. This model does not account for sampling bias of r^2 due to small population sizes. We extended the model to account for sample size according to Weir and Hill [75]:

$$N_e = \left(\frac{1}{4c}\right) \left(\frac{1}{r_{adj}^2} - 1\right)$$

where $r_{adj}^2 = r^2 - \frac{1}{(2*N)}$ with N being the sample size. To reduce the sampling error and bias in estimates of N_e , N_e was estimated only for populations that have a sample size ≥ 20 . The inclusion of related animals in the sample can cause an upwards bias of N_e (particularly recent N_e). Estimates of N_e presented in this paper are from samples that exclude animals that have a GRM relationship > 0.2 with one or more animals.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-020-07270-x>.

Additional file 1: Figure S1. PC1 vs. PC2 when using the whole dataset. (a) Expanded plot *Bos indicus* breeds section. (b) Expanded plot African *Bos taurus* breeds section.

Additional file 2: Figure S2. PCA using African *Bos taurus* populations (a) Plot of PC1 vs PC2. (b) PC1 vs PC3.

Additional file 3: Figure S3. Decline of linkage disequilibrium (r^2) with physical distance (kbp). (a) *Bos taurus* dairy breeds. (b) *Bos indicus* reference breeds. (c) East and South African indigenous breeds. (d) West African indigenous breeds.

Additional file 4: Figure S4. Effective population size over past generations using r^2 (log-scaled) for (a) *Bos taurus* dairy breeds, (b) *Bos indicus* breeds, (c) East and Southern African indigenous breeds, and (d) West African indigenous breeds.

Additional file 5: Figure S5. Effective population size over past generations using r_{adj}^2 (log-scaled) for (a) *Bos taurus* dairy breeds, (b) *Bos indicus* breeds, (c) East and Southern African indigenous breeds, and (d) West African indigenous breeds.

Additional file 6: Table S1. Estimates (\pm SD) of F_{IS} (diagonal) within the breeds and pairwise F_{ST} (above diagonal) values between the breeds.

Authors' contributions

NZG, JPG, and EMS conceived and designed the outline for the study. NZG performed the data analysis and drafted the manuscript. EMS and HA gave methodological support. NZG, JPG, EMS, HA, and KM interpreted the results. All authors read and approved the final manuscript.

Funding

NZG was funded by the University of New England International Postgraduate Research Award (UNE IPRA). The research was funded in part by the Bill and Melinda Gates Foundation and with UK aid from the UK Government's Department for International Development (Grant Agreement OPP1127286) under the auspices of the Centre for Tropical Livestock Genetics and Health (CTLGH), established jointly by the International Livestock Research Institute, the University of Edinburgh, and SRUC (Scotland's Rural College). The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the funding providers.

Availability of data and materials

Data were sourced from a variety of public domain and privately held databases as detailed in the paper. In most cases, the data held privately is available on request to the institution owning the data.

Ethics approval and consent to participate

Not applicable. Data sourced from previous studies.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Centre for Genetic Analysis and Applications, School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia.

²International Livestock Research Institute and Centre for Tropical Livestock Genetics and Health, Nairobi, Kenya.

Received: 2 June 2020 Accepted: 23 November 2020

Published online: 07 December 2020

References

- Gifford-Gonzalez D, Hanotte O. Domesticating animals in Africa: implications of genetic and archaeological findings. *J World Prehist.* 2011;24(1):1–23.
- Marshall F, Hildebrand E. Cattle before crops: the beginnings of food production in Africa. *J World Prehist.* 2002;16(2):99–143.
- Achilli A, Olivieri A, Pellecchia M, Ubaldi C, Colli L, Al-Zahery N, Accetturo M, Pala M, Kashani BH, Perego UA. Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Curr Biol.* 2008;18(4):R157–8.
- Verdugo MP, Mullin VE, Scheu A, Mattiangeli V, Daly KG, Delsler PM, Hare AJ, Burger J, Collins MJ, Kehati R. Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science.* 2019;365(6449):173–6.
- Stock F, Gifford-Gonzalez D. Genetics and African cattle domestication. *Afr Archaeol Rev.* 2013;30(1):51–72.
- Payne W, Hodges J. *Tropical cattle: origins, breeds and breeding policy.* Oxford: Blackwell Science; 1995.
- Smith AB. *African herders: emergence of pastoral traditions.* Walnut Creek, CA: AltaMira Press; 2005.
- Grigson C. An African origin for African cattle?—some archaeological evidence. *Afr Archaeol Rev.* 1991;9(1):119–44.
- Marshall F. Rethinking the role of *Bos indicus* in sub-Saharan Africa. *Curr Anthropol.* 1989;30(2):235–40.
- Rege J, Ayalew W, Getahun E, Hanotte O, Dessie T. Domestic animal genetic resources information system (DAGRIS). Addis Ababa: International Livestock Research Institute; 2007.
- Rege J, Tawah C. The state of African cattle genetic resources II. Geographical distribution, characteristics and uses of present-day breeds and strains. *Anim Genet Resour.* 1999;26:1–25.
- Lenstra J, Bradley D. Systematics and phylogeny of cattle. In: Fries R, Ruvinsky A, editors. *The genetics of cattle.* Wallingford: CAB International; 1999. p. 1–14.
- Weerasinghe MSPW. *Use of Genetic Polymorphisms to Assess the Genetic Structure and Breed Composition of Crossbred Animals: PhD Thesis, the University of New England;* 2014.
- Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P. Evidence for two independent domestications of cattle. *Proc Natl Acad Sci.* 1994;91(7):2757–61.
- Loftus R, Hugh DM, Ngere L, Balain D, Badi A, Bradley D, Cunningham E. Mitochondrial genetic variation in European, African and Indian cattle populations. *Anim Genet.* 1994;25(4):265–71.
- Bradley DG, MacHugh DE, Cunningham P, Loftus RT. Mitochondrial diversity and the origins of African and European cattle. *Proc Natl Acad Sci.* 1996;93(10):5131–5.
- Bradley D, MacHugh D, Loftus R, Sow R, Hoste C, Cunningham E. Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in west African trypanotolerant cattle populations. *Anim Genet.* 1994;25(5):7–12.
- Hanotte O, Tawah C, Bradley D, Okomo M, Verjee Y, Ochieng J, Rege J. Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds. *Mol Ecol.* 2000;9(4):387–96.
- MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics.* 1997;146(3):1071–86.
- Gebrehiwot NZ, Strucken EM, Aliloo H, Marshal K, Gibson JP. Genetic structure and differentiation among African *Bos taurus* cattle breeds. In: AAABG 2019: 2019: Association for the Advancement of Animal Breeding and Genetics; 2019.
- Marshall K, Tebug S, Salmon GR, Tapio M, Juga J, Missohou A. Improving dairy cattle productivity in Senegal, vol. 22. Nairobi: ILRI Policy Brief, ILRI; 2017.
- Chen N, Cai Y, Chen Q, Li R, Wang K, Huang Y, Hu S, Huang S, Zhang H, Zheng Z. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat Commun.* 2018;9(1):2337.
- Wu D-D, Ding X-D, Wang S, Wójcik JM, Zhang Y, Tokarska M, Li Y, Wang M-S, Faruque O, Nielsen R. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat Ecol Evol.* 2018;2(7):1139.
- Wang K, Lenstra JA, Liu L, Hu Q, Ma T, Qiu Q, Liu J. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Commun Biol.* 2018;1(1):169.
- Hiendleder S, Lewalski H, Janke A. Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication. *Cytogenet Genome Res.* 2008;120(1–2):150–6.
- Flori L, Moazami-Goudarzi K, Alary V, Araba A, Boujenane I, Boushaba N, Casabianca F, Casu S, Ciampolini R, Coeur D'Acier A. A genomic map of climate adaptation in Mediterranean cattle breeds. *Mol Ecol.* 2019;28(5):1009–29.
- Gautier M, Flori L, Riebler A, Jaffrézic F, Laloé D, Gut I, Moazami-Goudarzi K, Foulley J-L. A whole genome Bayesian scan for adaptive genetic divergence in west African cattle. *BMC Genomics.* 2009;10(1):550.
- Kim J, Hanotte O, Mwai OA, Dessie T, Bashir S, Diallo B, Agaba M, Kim K, Kwak W, Sung S. The genome landscape of indigenous African cattle. *Genome Biol.* 2017;18(1):34.
- Makina SO, Whitacre LK, Decker JE, Taylor JF, MacNeil MD, Scholtz MM, van Marle-Köster E, Muchadeyi FC, Makgahlela ML, Maiwashe A. Insight into the genetic composition of south African Sanga cattle using SNP data from cattle breeds worldwide. *Genet Sel Evol.* 2016;48(1):88.
- Park SD, Magee DA, McGettigan PA, Teasdale MD, Edwards CJ, Lohan AJ, Murphy A, Braud M, Donoghue MT, Liu Y. Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biol.* 2015;16(1):234.
- Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JEO. African pastoralism: genetic imprints of origins and migrations. *Science.* 2002;296(5566):336–9.
- Consortium BH. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science.* 2009;324(5926):528–32.
- Strucken EM, Al-Mamun HA, Esquivelzeta-Rabell C, Gondro C, Mwai OA, Gibson JP. Genetic tests for estimating dairy breed proportion and parentage assignment in east African crossbred cattle. *Genet Sel Evol.* 2017;49(1):67.
- Yurchenko AA, Daetwyler HD, Yudin N, Schnabel RD, Vander Jagt CJ, Soloshenko V, Lhasaranov B, Popov R, Taylor JF, Larkin DM. Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Sci Rep.* 2018;8(1):1–16.

35. Frischknecht M, Pausch H, Bapst B, Signer-Hasler H, Flury C, Garrick D, Stricker C, Fries R, Gredler-Grandl B. Highly accurate sequence imputation enables precise QTL mapping in Brown Swiss cattle. *BMC Genomics*. 2017; 18(1):999.
36. Sempéré G, Moazami-Goudarzi K, Eggen A, Laloë D, Gautier M, Flori L. WIDD E: a web-interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC Genomics*. 2015;16(1):1–8.
37. Pitt D, Sevane N, Nicolazzi EL, MacHugh DE, Park SD, Colli L, Martinez R, Bruford MW, Orozco-terWengel P. Domestication of cattle: two or three events? *Evol Appl*. 2019;12(1):123–36.
38. Ndiaye NP, Sow A, Dayo G-K, Ndiaye S, Sawadogo GJ, Sembène M. Genetic diversity and phylogenetic relationships in local cattle breeds of Senegal based on autosomal microsatellite markers. *Vet World*. 2015; 8(8):994.
39. Marshall K, Salmon G, Tebug S, Juga J, MacLeod M, Poole J, Baltenweck I, Missohou A. Net benefits of smallholder dairy cattle farms in Senegal can be significantly increased through the use of better dairy cattle breeds and improved management practices. *J Dairy Sci*. 2020;103(9):8197–217.
40. Decker JE, McKay SD, Rolf MM, Kim J, Alcalá AM, Sonstegard TS, Hanotte O, Götherström A, Seabury CM, Praharani L. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet*. 2014;10(3): e1004254.
41. Mastrangelo S, Tolone M, Jemaa SB, Sottile G, Di Gerlando R, Cortés O, Senczuk G, Portolano B, Pilla F, Ciani E. Refining the genetic structure and relationships of European cattle breeds through meta-analysis of worldwide genomic SNP data, focusing on Italian cattle. *Sci Rep*. 2020;10(1):1–13.
42. Grema M, Traoré A, Issa M, Hamani M, Abdou M, Soudré A, Sanou M, Pichler R, Tamboura HH, Alhassane Y. Short tandem repeat (STR) based genetic diversity and relationship of indigenous Niger cattle. *Arch Anim Breed*. 2017; 60(4):399–408.
43. Macfarlane J. Evolutionary development of dual-purpose zebu cattle (the Mpwapwa breed project). In: *A review of livestock research in Tanzania 1950-1970*; 1971.
44. Gautier M, Naves M. Footprints of selection in the ancestral admixture of a New World creole cattle breed. *Mol Ecol*. 2011;20(15):3128–43.
45. Kibiego MB, Lagat JK, Bebe BO. Competitiveness of smallholder Milk production Systems in Uasin Gishu County of Kenya. *J Econ Sust Dev*. 2015; 6(10):39–46.
46. Balikowa D. Dairy development in Uganda. A review of Uganda's dairy industry. Dairy Dev Authority Uganda. 2011;3202(1):1–215.
47. EthiopianBiodiversityInstitute: Ethiopia's revised national biodiversity strategy and action plan. Government Report. 2014.
48. Cheruiyot EK, Bett RC, Amimo JO, Yi Z, Mrode R, Mujibi FD. Signatures of selection in admixed dairy cattle in Tanzania. *Front Genet*. 2018;9:607.
49. Ema PN, Lassila L, Missohou A, Marshall K, Tapio M, Tebug SF, Juga J. Milk production traits among indigenous and crossbred dairy cattle in Senegal. *Afr J Food Agric Nutr Dev*. 2018;18(3):13572–87.
50. Ardlie KG, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet*. 2002;3(4):299–309.
51. Qanbari S, Pimentel E, Tetens J, Thaller G, Lichtner P, Sharifi A, Simianer H. The pattern of linkage disequilibrium in German Holstein cattle. *Anim Genet*. 2010;41(4):346–56.
52. Hayes BJ, Visscher PM, McPartlan HC, Goddard ME. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res*. 2003;13(4):635–43.
53. Decker JE, Vasco DA, McKay SD, McClure MC, Rolf MM, Kim J, Northcutt SL, Bauca S, Woodward BW, Schnabel RD. A novel analytical method, birth date selection mapping, detects response of the Angus (*Bos taurus*) genome to selection on complex traits. *BMC Genomics*. 2012;13(1):606.
54. Clutton-Brock J. *A natural history of domesticated mammals*: Cambridge University Press; 1999.
55. Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature*. 2002;418(6898):700.
56. Meuwissen T. Genetic management of small populations: a review. *Acta Agric Scand Sect A*. 2009;59(2):71–9.
57. Phillipson DW. *African archaeology*: Cambridge University Press; 2005.
58. Hill WG. Estimation of effective population size from data on linkage disequilibrium. *Genet Res*. 1981;38(3):209–16.
59. Corbin LJ, Liu A, Bishop S, Woolliams J. Estimation of historical effective population size using linkage disequilibria with marker data. *J Anim Breed Genet*. 2012;129(4):257–70.
60. Ema P, Missohou A, Marshall K, Tebug S, Juga J, Tapio M. Genetic admixture and identity by descent in Senegalese dairy cattle. In: *In Proceedings 36th International Society for Animal Genetics (ISAG) Dublin, Ireland: 2017*, vol. 53; 2017. p. 54.
61. Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007;23(10):1294–6.
62. R Core Team: R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria. 2018.
63. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91(11):4414–23.
64. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655–64.
65. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*. 2011;12(1):246.
66. Weerasinghe S, Gibson J, Gondro C, Jeyaruban G. Use of genetic polymorphisms to assess the genetic structure and breed composition of crossbred animals; 2016.
67. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984;38(6):1358–70.
68. Nei M. F-statistics and analysis of gene diversity in subdivided populations. *Ann Hum Genet*. 1977;41(2):225–33.
69. Kolde R. pheatmap: Pretty Heatmaps. R package version 1.0. 8; 2015.
70. Loh P-R, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, Schoenherr S, Forer L, McCarthy S, Abecasis GR. Reference-based phasing using the haplotype reference Consortium panel. *Nat Genet*. 2016;48(11): 1443.
71. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
72. Hill W, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet*. 1968;38(6):226–31.
73. McVean G. Linkage disequilibrium, recombination and selection. *Handbook of statistical genetics*; 2007. p. 909–44.
74. Sved J. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol*. 1971;2(2):125–41.
75. Weir B, Hill WG. Effect of mating structure on variation in linkage disequilibrium. *Genetics*. 1980;95(2):477–88.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

