



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya

Citation for published version:

Mbole-Kariuki, MN, Sonstegard, T, Orth, A, Thumbi, SM, Bronsvoort, M, Kiara, H, Toye, P, Conradie, I, Jennings, A, Coetzer, K, Woolhouse, MEJ, Hanotte, O & Tapio, M 2014, 'Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya' *Heredity*, vol. 113, no. 4, pp. 297-305. DOI: 10.1038/hdy.2014.31

Digital Object Identifier (DOI):

[10.1038/hdy.2014.31](https://doi.org/10.1038/hdy.2014.31)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Heredity

Publisher Rights Statement:

This work is licensed under a Creative Commons Attribution 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



ORIGINAL ARTICLE

Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya

MN Mbole-Kariuki¹, T Sonstegard², A Orth³, SM Thumbi⁴, BMdeC Bronsvoort⁵, H Kiara³, P Toye³, I Conradie⁶, A Jennings⁵, K Coetzer⁶, MEJ Woolhouse⁴, O Hanotte¹ and M Tapio⁷

The Kenyan East African zebu cattle are valuable and widely used genetic resources. Previous studies using microsatellite loci revealed the complex history of these populations with the presence of taurine and zebu genetic backgrounds. Here, we estimate at genome-wide level the genetic composition and population structure of the East African Shorthorn Zebu (EASZ) of western Kenya. A total of 548 EASZ from 20 sub-locations were genotyped using the Illumina BovineSNP50 v. 1 beadchip. STRUCTURE analysis reveals admixture with Asian zebu, African and European taurine cattle. The EASZ were separated into three categories: substantial ($\geq 12.5\%$), moderate ($1.56\% < X < 12.5\%$) and non-introgressed ($\leq 1.56\%$) according to the European taurine genetic proportion. The non-European taurine introgressed animals ($n = 425$) show an unfluctuating zebu and taurine ancestry of 0.84 ± 0.009 s.d. and 0.16 ± 0.009 s.d., respectively, with significant differences in African taurine (AT) and Asian zebu backgrounds across chromosomes ($P < 0.0001$). In contrast, no such differences are observed for the European taurine ancestry ($P = 0.1357$). Excluding European introgressed animals, low and nonsignificant genetic differentiation and isolation by distance are observed among sub-locations ($F_{st} = 0.0033$, $P = 0.09$; $r = 0.155$, $P = 0.07$). Following a short population expansion, a major reduction in effective population size (N_e) is observed from approximately 240 years ago to present time. Our results support ancient zebu \times AT admixture in the EASZ population, subsequently shaped by selection and/or genetic drift, followed by a more recent exotic European cattle introgression.

Heredity advance online publication, 16 April 2014; doi:10.1038/hdy.2014.31

INTRODUCTION

The East African cattle group is a valuable genetic resource with a complex origin. The first African cattle were of taurine type *Bos taurus* (Gifford-Gonzalez and Hanotte, 2011). According to latest mitochondrial DNA results they originated from the geographic center of cattle domestication in the Near East and separated from the other taurine types approximately 7000 years ago (Bonfiglio *et al.*, 2012). These taurine cattle entered Africa through its North-Eastern part via present day Egypt (Epstein, 1971; Blench and MacDonald, 2000; Gifford-Gonzalez and Hanotte, 2011). Zebu cattle (*Bos indicus*) originated in the Indian subcontinent and migrated into Africa more recently (Gifford-Gonzalez and Hanotte, 2011). The earliest undisputed evidences of zebu cattle dated from the first mid-millennium AD (Gifford-Gonzalez and Hanotte, 2011). They may have subsequently penetrated Africa in two waves (Hanotte *et al.*, 2002), with the second wave possibly facilitated by the rinderpest epidemic (Blench, 1993; Paynes and Hodges, 1997). Contemporary cattle from the eastern part of Africa are predominantly phenotypically classified

as zebu (Rege and Tawah, 1999; DAGRIS, 2007). As to whether or not the African aurochs *B. primigenius africanus* (now extinct) contributed to the genetic stock of African domestic cattle remains unknown (Gifford-Gonzalez and Hanotte, 2011).

Currently, Africa is home to over 150 recognized cattle breeds that comprise of a mosaic of zebu, taurine and crossbreeds (indicine and taurine), the latter sometimes referred to as sanga (Rege and Tawah, 1999). In Kenya, owing to the effect of tribal boundaries and socioeconomic cultures, different strains of the East African Shorthorn Zebu (EASZ) are recognized (Rege, 1999). These include the Kavirondo zebu reared by the Luo and Luhya communities, and the Teso zebu reared by the Teso community who mainly inhabit western Kenya (DAGRIS, 2007). Genetic studies carried out using microsatellite markers (Rege *et al.*, 2001) show that Kenyan zebu populations are zebu–taurine hybrids with a major zebu genetic component. Further studies with Y-chromosomal markers (Hanotte *et al.*, 2000) and mitochondrial DNA markers (Bradley *et al.*, 1996) are in agreement with a male-mediated zebu introgression of the taurine animals.

¹Ecology and Evolution, School of Life Sciences, University of Nottingham, Nottingham, UK; ²United States Department of Agriculture, Agriculture, Agricultural Research Services, Bovine Functional Genomics Laboratory, Beltsville, MD, USA; ³International Livestock Research Institute, Nairobi, Kenya; ⁴Centre for Immunity, Infection and Evolution, Ashworth Laboratories, University of Edinburgh, Edinburgh, UK; ⁵The Roslin Institute, Easter Bush, University of Edinburgh, Edinburgh, UK; ⁶Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Onderstepoort, South Africa and ⁷MTT FI-31600, Jokioinen, Finland

Correspondence: Current address: Dr MN Mbole-Kariuki, African Union Interafrican Bureau for Animal Resources (AU-IBAR), Kenindia Business Park Building, Museum Hill, Westlands Road, PO BOX 30786, 00100 Nairobi, Kenya.
E-mail: mary.mbole-kariuki@au-ibar.org or ndila.kariuki@gmail.com
or Professor O Hanotte, Ecology and Evolution, School of Life Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK.
E-mail: olivier.hanotte@nottingham.ac.uk

Received 24 July 2013; revised 11 February 2014; accepted 20 February 2014

In more recent times, a wave of 'exotic' taurine cattle introductions has set off following governmental and international livestock development agendas aimed at improving livestock productivity (Hanotte *et al.*, 2002). These cattle are of European taurine genetic backgrounds and include breeds such as Holstein–Friesian and Jersey as reported in Kenya (Weir *et al.*, 2009) and Ethiopia (Haile *et al.*, 2011). The rapidly changing socioeconomic and cultural environment of rural farming is favoring these introductions. This shift of focus is to a perceived economically beneficial animal as opposed to an ecologically fit one, leading to active breed replacement or cross-breeding programs within Eastern Africa, particularly in the highlands and peri-urban areas (Bebe *et al.*, 2003).

Today, indigenous zebu cattle are the commonest cattle type across most parts of Eastern Africa (Rege and Tawah, 1999). Their wide distribution is a possible consequence of their environmental adaptive traits not found in exotic taurine breeds. These include resistance or tolerance to tropical diseases and their vectors (Latif *et al.*, 1991a,b; Mattioli *et al.*, 1993; Lawrence *et al.*, 1996; Wambura *et al.*, 1998; Hanotte *et al.*, 2003), survival on poor quality forages/pastures (Bonsma, 1973; Turner, 1980) and thermal stress tolerance (Carvalho *et al.*, 1995; Hammond *et al.*, 1996; Gaughan *et al.*, 1999; Hansen, 2004). In addition, African zebu cattle are still a reliable source of draught power (Rege *et al.*, 2001).

African taurine (AT) are on the verge of extinction in East Africa (Rege, 1999), but their environmental genetic adaptation (for example, infectious disease tolerance) may have survived in East African zebu and their crossbreeds. However, despite being well adapted to the harsh tropical environment, African zebu and taurine cattle remain poor producers in comparison with exotic breeds raised in the temperate environments (Rege *et al.*, 2001). A medium to long-term solution for sustainable improvement of productivity in the tropics may be to combine ecologically adaptive zebu traits and economically important exotic cattle traits in crossbreed animals.

The recent availability of genome-wide scan tools are offering new opportunities for genetic characterization (Gautier *et al.*, 2009, 2010; McTavish *et al.*, 2013), genome-wide association studies (Van Tassel *et al.*, 2008; Matukumalli *et al.*, 2009; Settles *et al.*, 2009; Pant *et al.*, 2010), the detection of signatures of selection for productivity (Barendse *et al.*, 2009) and genomic evaluation (Wiggins *et al.*, 2008, 2010) of cattle populations. Most of these studies have been carried out predominantly on European dairy and beef cattle breeds (The Bovine HapMap Consortium, 2009; Melka and Schenkel, 2012) with the exception of Gautier *et al.* (2009) and Gautier and Naves (2011) who worked on West African and Caribbean Creole cattle, respectively.

In this study, we characterize at genome-wide level the genetic diversity and architecture of 548 EASZ, an indigenous cattle population from Kenya. We present evidences for ancient zebu × taurine admixture, population bottleneck and expansion as well as the presence of recent and ongoing exotic taurine introgression. No significant genetic differentiation in non-European introgressed animals is observed across the studied sites. Moreover, our study further provides insight in the usefulness and limitations of low-density single-nucleotide polymorphisms (SNPs) chips toward understanding the genomic architecture and history of indigenous African tropical cattle.

MATERIALS AND METHODS

Study cohort and sampling site

A total of 548 calves were sampled from 20 different randomly selected sub-locations that traverse four distinct ecological zones in Western and Nyanza provinces of Kenya (Supplementary Figure S1). Upon recruitment, blood samples were drawn from the jugular vein using a 10 ml sterile syringe. Five ml

of blood was mixed in sodium EDTA tubes in a 1:1 ratio with 'magic buffer' (which acted as an anti-coagulant, anti-fungus, anti-bacterial and preservative—Biogen Diagnostica, Villaviciosa De Odon, Spain). The tubes were labeled with their respective bar-coded tags before being stored at 4 °C at the International Livestock Research Institute—ILRI (Nairobi, Kenya) biobank. DNA was extracted using the Nucleon Genomic DNA extraction kit (Tepnel Life Sciences, Manchester, UK).

Genotyping and quality control

The Illumina BovineSNP50 v. 1 beadchip (Illumina Inc., San Diego, CA, USA) includes 56 947 SNPs comprising 54 436 autosomal SNPs, 1341X chromosome SNPs and 1170 unmapped SNPs (the mapping of the genomic positions was done using the University of Maryland genome assembly v 3.0; www.cbcb.umd.edu/research/bos_taurus_assembly.shtml). Genotyping of the 548 calves was carried out at the USDA-ARS bovine functional (Beltsville, MD, USA) and GeneSeek (<http://www.neogen.com/geneseek/>) laboratories. An additional 158 reference animals representing two European taurine breeds (Holstein $n=64$ and Jersey $n=28$), one AT breed (N'dama $n=25$), one Asian zebu breed (Nelore $n=21$) and one East African admixed breed (Sheko $n=20$, Ethiopia) breeds were drawn from the Bovine HapMap Consortium (2009).

Quality control was carried out using *GenABEL* program (Aulchenko *et al.*, 2007) in R (R Development Core Team, 2009). The *check.marker* function was used to prune individual calves that failed to pass the inclusion criterion of successful genotypes per calf of >90% and an average identical by state threshold of >90% (excluding both animals).

Only mapped autosomal SNPs ($n=54\,436$) were screened. Autosomal SNPs with call rates of <90% were excluded. Unless stated, no minor allele frequency threshold (to avoid exclusion of SNPs that maybe informative in a single breed or population) and Hardy and Weinberg Equilibrium (as introgression may result in loci not in Hardy and Weinberg equilibrium) criteria, were applied. A total of 6151 SNPs failed the inclusion criterion, leaving a total 48 285 SNP for analysis. A random subset of 45 000 SNPs was used for STRUCTURE, principal component analysis, genetic relatedness and genetic differentiation analyses. No random sampling of SNPs was performed for linkage disequilibrium (LD) and effective population size calculations. Genotyping data has been deposited in Dryad (Murray *et al.*, 2013a).

Admixture analysis

The extent of admixture and the origin of the different genetic proportions were investigated, using a Bayesian clustering method implemented in the STRUCTURE program (Pritchard *et al.*, 2000, 2007). Five independent replicates of an admixed model with independent allele frequencies were run for a burn-in period of 50 000 iterations and 100 000 Markov Chain Monte Carlo steps for $K=1$ to $K=7$. The mean output files from CLUSTER Matching and Permutation Program—CLUMPP v 1.1.2 (Jakobsson and Rosenberg, 2007) were used as input files to graphically display the population structure barplots using the *barplot* function in R (R Development Core Team, 2009). The population structure analyses using 45K markers were run on three calf sample sets: (i) full study population ($n=548$) and reference breeds ($n=158$) totaling 706 animals (data set 1), (ii) non-European introgressed calves ($n=425$) and reference breeds ($n=158$) giving a total of 583 animals (data set 2) and (iii) non-European introgressed calves ($n=425$) (data set 3). The Ward clustering method (Ward, 1963) using the *hclust* function in R (R Development Core Team, 2009) was used to identify discontinuities in the distribution of the European taurine ancestry within data set 1. Multiple methods were used to evaluate the optimal number of genome backgrounds within the study population (Falush *et al.*, 2003; Evanno *et al.*, 2005; Pritchard *et al.*, 2007).

Principal component analysis

Principal component analyses were applied on the full study population and reference breeds (data set 1) and the non-European introgressed calves (data set 3) using *adegenet* v 1.3.1 (Jombart, 2008) to read the data files into R and *ade4* genetic package to calculate the principal components and eigenvalues (Dray and Dufour, 2007). Both packages are found in R (R Development Core Team, 2009).

Genetic relatedness and genetic differentiation

Global and pairwise F_{st} statistics were calculated using the R-based *hierfstat* 0.04–6 package (Goudet, 2005). The assessment of the influence of physical geographic distance on genetic distance between calf pairs within and between the sub-locations (data set 3) was tested with the Mantel test. The analysis was carried out with *adegenet* v 1.3.1 (Jombart, 2008) genetic package in R (R Development Core Team, 2009). The geographic coordinates were converted to kilometers using conversion units based on the World Geodetic System 1984 (WGS84) spheroid. The functions *dist* and *mantel.randtest* were used to calculate the pairwise geographic distances and the Mantel test statistic, respectively. The pairwise rescaled F_{st} estimates ($F_{st}/(1-F_{st})$) were used for the estimation of genetic differentiation between sub-locations in relation to geographic distance (Rousset, 1997).

LD and effective population size in pure EASZ

Pairwise LD, measured as squared correlation coefficient r^2 (Hill and Robertson, 1968), was calculated in *GenABEL* (Aulchenko *et al.*, 2007) for the non-introgressed European calves (data set 3). The decrease of LD as a function of distance between markers was evaluated for pairs up to 4 Mb apart using markers with minor allele frequency above 0.01. The expected value was predicted with Loess local regression of second degree using the 5% of the data closest to the estimation point in the local regression as implemented in R (R Development Core Team, 2009). For estimating the effective population size, 11 bins (<0.1, 0.1–0.2, 0.2–0.3, 0.3–0.5, 0.5–1.0, 1.0–1.5, 1.5–2.0, 2.0–2.5, 2.5–3.0, 3.0–3.5 and >3.5 Mb) were used, and the mean for each bin used to obtain the expected r^2 .

Estimation of the ancestral effective population size was calculated using the Weir and Hill (1980) adjusted formula, $E(r^2) = [1/(1+4N_e c)] + (1/n)$. Where $E(r^2)$ is the expected LD, N_e is the effective population size, c is the recombination frequency, n is the chromosome size or twice the sample size, c was calculated using the Haldane's mapping function (Haldane, 1919), with the average marker distance between adjacent SNPs expressed in Morgans, assuming 1 Mb = 1 cM (de Roos *et al.*, 2008). The estimate for a bin relates to the N_e for the generation $t = 1/2c$ (Hayes *et al.*, 2003) counted backwards from the genotyped generation. A generation length of 6 years was assumed (Mahadevan, 1955).

Performance and ascertainment bias of the Illumina BovineSNP50 v.1 beadchip in Kenyan EASZ

To assess the Illumina BovineSNP50 v.1 beadchip performance as a tool of estimating zebu and taurine admixture proportions. STRUCTURE analyses at $K=3$ were performed using data sets of randomly selected sets of SNPs ($n=5000, 10000, 15000, 25000, 35000$ and 45000) drawn from the cleaned data set of 48285 markers. These particular analyses were run using the moderate and substantial calf categories consisting of 123 calves and 158 animals representing the five reference breeds (total number 281). Ancestral genome proportions were generated from the STRUCTURE runs and used in the subsequent linear correlation analyses carried out between markers sets in R (R Development Core Team, 2009).

RESULTS

Out of 48285 autosomal SNPs remaining in the data set following quality control, 11269 markers were monomorphic across the EASZ population. The mean observed heterozygosity (H_o) within EASZ calves ($n=548$) and reference breeds are indicated in Table 1. EASZ population shows an average H_o of 0.25 ± 0.02 s.d. with no significant differences across sub-locations ($P > 0.001$).

STRUCTURE runs from $K=2$ to $K=5$ using 45000 random SNPs are presented in Figure 1. Applying the Evanno *et al.* (2005) method suggests $K=2$ as the optimal partition (Supplementary Figure S2), which is the uppermost relevant hierarchy reflecting the taurine and indicine cattle split. A clear but less drastic improvement in the fit of the model is visible by increasing K to 3, revealing previously documented findings that highlight a further taurine split to African and European taurines (The Bovine HapMap Consortium, 2009).

Table 1 Heterozygosity estimates of EASZ and four reference breeds

Breed	Observed heterozygosity (H_o)
EASZ	0.25 ± 0.02 s.d.
Holstein–Friesian	0.33 ± 0.01 s.d.
Jersey	0.25 ± 0.03 s.d.
N'dama	0.17 ± 0.08 s.d.
Ethiopian Sheko	0.26 ± 0.003 s.d.

Abbreviation: EASZ, East African Shorthorn Zebu.

The increase from 3 to 4 genetic clusters is minimal and does not lead to a new individual breed cluster. However, a notable improvement is observed when increasing K to 5, revealing a finer resolution separating the two European taurine breeds (Supplementary Figure S2). Above $K=5$, the increase in goodness of fit with larger K values are only incremental, suggesting that they do not reveal significant phylogenetic structure (Falush *et al.*, 2003, Prichard *et al.*, 2007).

STRUCTURE results of $K=3$ are in agreement with prior information about the main genetic architecture of the cattle on the African continent being of three different ancestries using micro-satellite loci; Asian zebu *B. indicus*, AT *B. taurus* and European taurine *B. taurus* (Hanotte *et al.*, 2002; The Bovine HapMap Consortium, 2009). Of interest, a subset of the EASZ calves has European taurine ancestry, while the other East African cattle breed studied, the Sheko, shows no European taurine introgression (Supplementary Table S1). In addition, at $K=3$, the Jersey breed presents a shared genetic background (mean ancestral proportion -0.12 , Supplementary Table S1) with the N'dama breed. However, this is not observed at $K=5$ (Figure 1). Possible AT membership is also observed in some Holstein–Friesian animals up to a proportion of 0.10–0.11 ($n=3$) (Supplementary Table S1). Principal component analysis on the same data set shows that PC1, explaining 65% of the variation, separates the indicine and taurine breeds; whereas PC2, explaining 14% of the variation, separates the AT breed (represented here by the N'dama of West Africa) from the European breeds (Figure 2). For $K=4$, the proportion of European taurine in EASZ remains the same, but not the inferred ancestral proportion of zebu and AT background (Figure 1). In addition, there is hardly any genetic background shared between the EASZ and the West African cattle (N'dama). However, a substantial proportion of the EASZ genome remains shared with the Nelore (Asian zebu). The largest proportion of genome ancestry present in EASZ is now nearly unique to the EASZ and Sheko with only traces of it found within the Nelore (Figure 1). $K=5$ divides the inferred European ancestry between the two European breeds (Jersey and Holstein–Friesian) (Figure 1).

Using the Ward clustering method (Ward, 1963), we further analyzed the proportion of European taurine background in the EASZ (Figure 3a) based on the $K=3$ model. Based on three observed clusters, we defined three categories of calves: calves with $\geq 12.5\%$ European taurine background (category 1 representing animals with 'substantial' European introgression $n=29$), calves with between 1.56 and 12.5% European taurine introgression (category 2 representing the 'moderate' European taurine introgressed sample set $n=94$) and calves with $\leq 1.56\%$ European taurine background (category 3 representing the 'non-European' introgressed sample set $n=425$). The 123 introgressed calves, representing the moderate and substantial categories, are found within 12 sub-locations in the northern and central regions of the study area (Supplementary Table S1). A geospatial analysis of the substantial category reveals two main hotspots of European taurine introgression (Figure 3b), whereas the

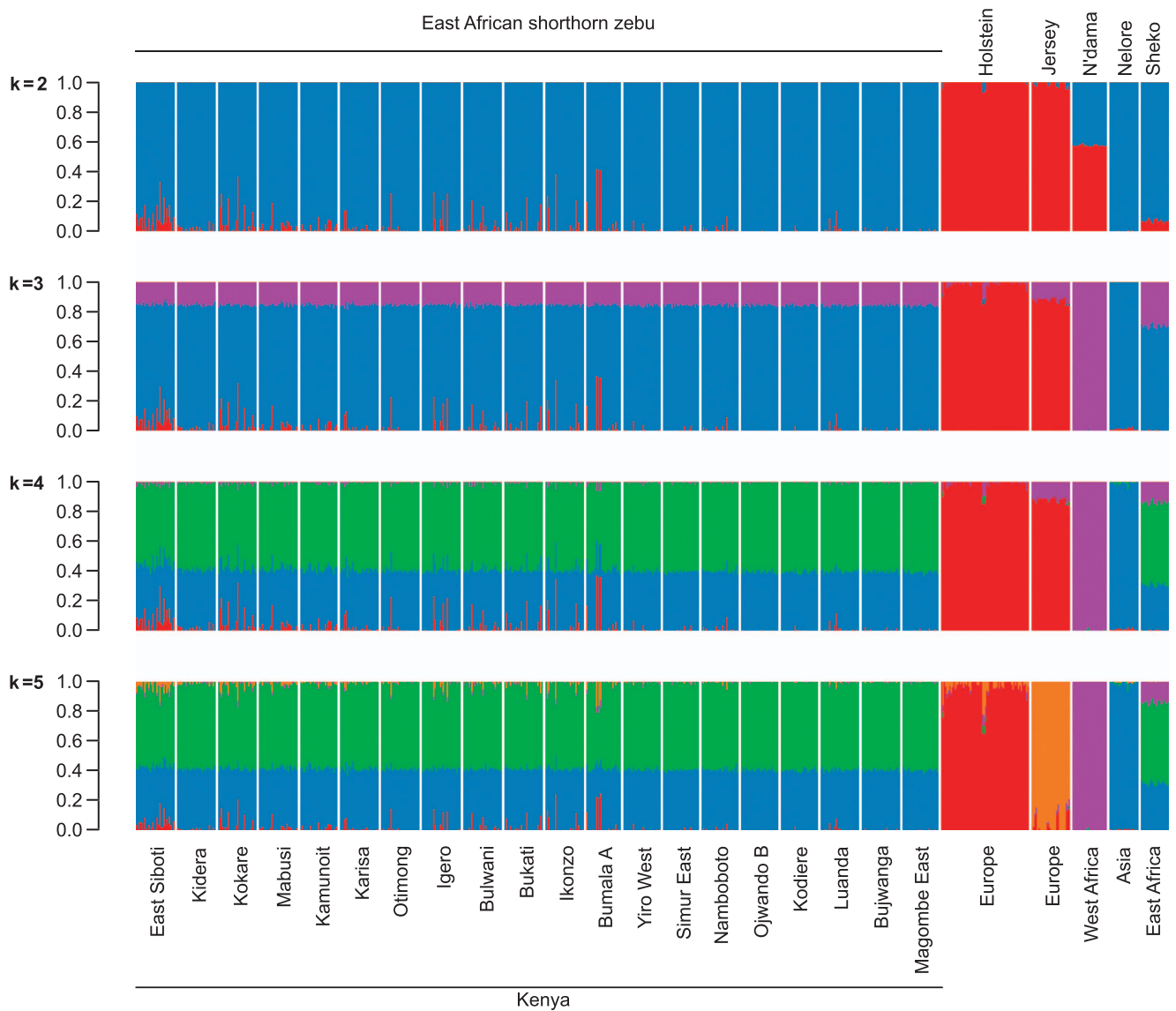


Figure 1 STRUCTURE bar plots of genetic membership proportions ($K=2$ to $K=5$). Each animal is represented by a vertical line divided into K colors. Breed names and locations are indicated at the top and bottom of the bar plots, respectively.

moderate category shows a north to south decrease of European taurine introgression ($r=0.82$, $P<0.0001$; Figure 3b).

STRUCTURE analysis ($K=3$) of data set 2 reveals a homogenous admixed EASZ population (Supplementary Figure S3) with average genetic proportions of 0.84 ± 0.009 s.d. and 0.16 ± 0.009 s.d. of AT and Asian zebu ancestries, respectively. Significant genome-wide difference of AT or zebu ancestry across calves ($P<0.0001$) was also observed.

Interestingly, the chromosome-wise analyses reveal that some of the 'non-introgressed' calves (Supplementary Figure S4a) have moderate proportion of European taurine ancestry on a small subset of chromosomes, similarly within the Sheko population. There are highly significant differences ($P<0.0001$, Supplementary Figure S4b and Supplementary Table S2) among chromosomes in the amount of AT ancestry in both introgressed and non-introgressed EASZ calves. However, for the same genetic ancestry (AT) in the Sheko population, no significant differences were observed ($P>0.05$; Supplementary Table S2). For the European taurine and Asian zebu ancestries, highly

significant differences between chromosomes are observed in the non-introgressed calves' cohort unlike in the Sheko population ($P>0.05$; Supplementary Table S2), as illustrated in Supplementary Figures S4a and c, respectively. Differences among calves are highly significant for European taurine ancestry in both introgressed and non-introgressed calves, as well as for Asian zebu ancestry in introgressed calves (Supplementary Figures S4a and c). In contrast, there are no differences in the AT ancestry among all the calves' cohorts and Sheko population (Supplementary Figure S4b).

Strong positive correlations for the AT and Asian zebu ancestry between EASZ and Sheko are observed across chromosomes ($r=0.89$, $P<0.0001$ and $r=0.94$, $P<0.0001$). Considering only EASZ calves with moderate and substantial European taurine introgression, no significant difference is observed in European taurine ancestry proportion across chromosomes ($P=0.136$).

Data set 3 allows addressing the possible presence of sub-structuring within the non-European taurine introgressed EASZ population. PC1 and PC2 explain 4% and 3% of the percentage of the total

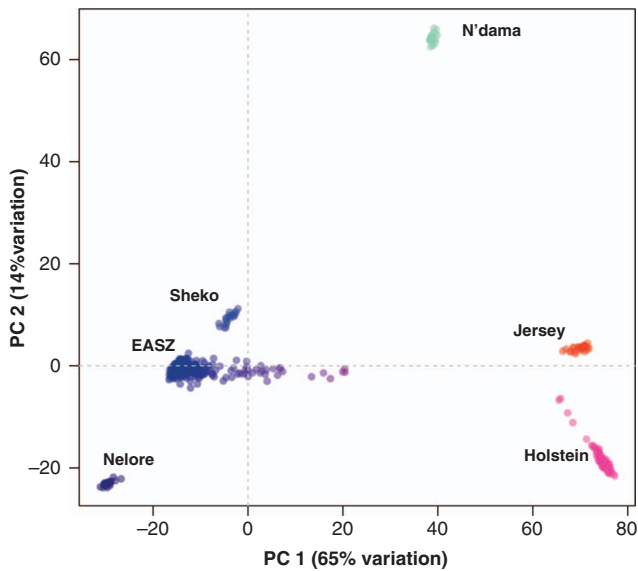


Figure 2 Principal component analysis plot of EASZ and reference breeds. PC1 and PC2 explain 65% and 14% of the total variance, respectively.

variance, respectively (Supplementary Figure S5a). Three genetic clusters are observed of which the largest includes 415 calves, while the second and third clusters include 4 and 6 calves from Bukati and Luanda sub-locations, respectively (Supplementary Figure S5a). This result is further illustrated with the STRUCTURE results ($K=3$), which distinguishes the same calves (Supplementary Figure S5b).

No genetic differentiation is observed between sub-locations ($F_{st}=0.0033$, $P=0.09$) after removing the moderate and substantial European introgressed calves. Mantel test between pairwise sub-locations genetic differentiation (calculated as $(F_{st}/(1-F_{st}))$ and geographic distances was not significant ($r=0.155$, $P=0.07$) for the pure calves category (Supplementary Figure S6).

The extent of LD over genome regions is related to effective population size. In the pure EASZ population, the LD decreased quickly (Supplementary Figure S7). The expected value was estimated with Loess curve based on a total of 2 368 859 comparisons between markers. At the minimum distance, the expected r^2 is 0.32. It decreases to 0.2 at 55.4 kb, to 0.1 at 200.3 kb and to 0.05 at 486.0 kb distance. LD begins to asymptote at the background level of $r^2=0.013$ around 2 MB. The LD results were used to estimate the effective population size.

The effective population size is characterized by a population decline over time (Figure 4). However, a short stint of increasing population size, starting around 126 generations ago (~ 756 years ago), is observed before a drastic population decrease starting around 40 generations ago (~ 240 years ago) and continuing to present day (Figure 4).

We assessed the effect of the number of markers for the estimation of genome ancestry in our population assuming three ancestral populations (indigenous AT, European taurine and Asian zebu) and using only calves with moderate and substantial European taurine introgression. We use linear correlation between sets of randomly selected number of markers using STRUCTURE's output of inferred genetic membership proportions at $K=3$. The analyses indicate that the estimation of the proportion of genome ancestry varied with the number of markers included, and that it depends on the genetic background being estimated (Supplementary Figures S8a–c).

As expected, the more markers the better the correlation. More particularly, a correlation coefficient of $r=0.99$ ($P<0.0001$) was obtained for the 15–25K pairwise marker comparison for the European background (Supplementary Figure S8a), and for the 25–35K pairwise marker comparison for the Asian zebu ancestry (Supplementary Figure S8b). However, for the estimation of the AT ancestral background a correlation coefficient of only $r=0.92$ ($P<0.0001$) was obtained for the 35–45K pairwise marker comparisons (Supplementary Figure S8c).

Pruning neighboring markers in strong LD (r^2 over 0.1–0.5) has been suggested as a way to reduce data redundancy in STRUCTURE analysis and ascertainment bias in the estimation of diversity. Approximately 44% of the neighboring marker pairs linkage were above the lower limit ($r^2=0.1$) and only 11% above the higher limit ($r^2=0.5$) within the non-introgressed calves cohort. The proportion of marker pairs in LD approximates the genome coverage of the chip, for example, in mapping applications. Taking into account the marker gaps, these proportions equal to coverage of 40 and 8%, with the respective r^2 limits. The median for neighboring marker pairs was 0.075. Therefore, no LD-based pruning was undertaken.

DISCUSSION

This study aimed at unraveling the population history and genetic structure of an indigenous population of EASZ of western Kenya at genome-wide level. At $K=3$, STRUCTURE analysis agrees with the principal component analysis dimension results indicating the presence of three genetic backgrounds. This is not surprising given our current understanding of the origin and history of these populations and it is in agreement with previous finding using microsatellite loci (Hanotte *et al.*, 2002). Moreover, using 45 000 randomly selected autosomal genome-wide SNP markers, we are able to unravel finer details of the extent of genome admixture (AT, European taurine and Asian zebu) within the studied populations (Figures 1 and 2). Assessment of the level of inferred ancestral proportion difference between animals (two EASZ calf cohorts and the Sheko breed) and autosomes (Supplementary Figures S4a–c) is also presented.

Our results support ancient zebu \times AT admixture in the EASZ population, subsequently shaped by selection and/or genetic drift, followed by a more recent exotic European cattle introgression. Indeed, we do observe very little variation among animals for the inferred zebu and AT ancestral proportions, at the contrary of the inferred European taurine background. It indicates that the AT and zebu genome ancestries have had time to 'diffuse' homogeneously among the calves of our study population, while the level of European taurine introgression, of more recent origin, still needs to reach an equilibrium (Figure 1). Interestingly, we observe differences in zebu or AT ancestry proportions among chromosomes, differences shared between the EASZ and the Sheko as revealed by the correlation analyses. It suggests at least a partial role of selection in shaping the genome architecture of present day indigenous East African cattle populations rather than only genetic drift. The EASZ and Sheko admixed (taurine \times zebu) populations occupy to some extent similar agro-ecological environments in different geographical locations (DAGRIS, 2007). It remains, however, unknown which common selection pressures may have shaped the genome of these crossbred populations. These effects may be attributed to environmental factors (for example, common infectious disease challenges) and/or may be the consequence of within genome selection pressures following the crossbreeding of cattle belonging to two distinct lineages that separated more than half a million years ago (Loftus *et al.*, 1994;

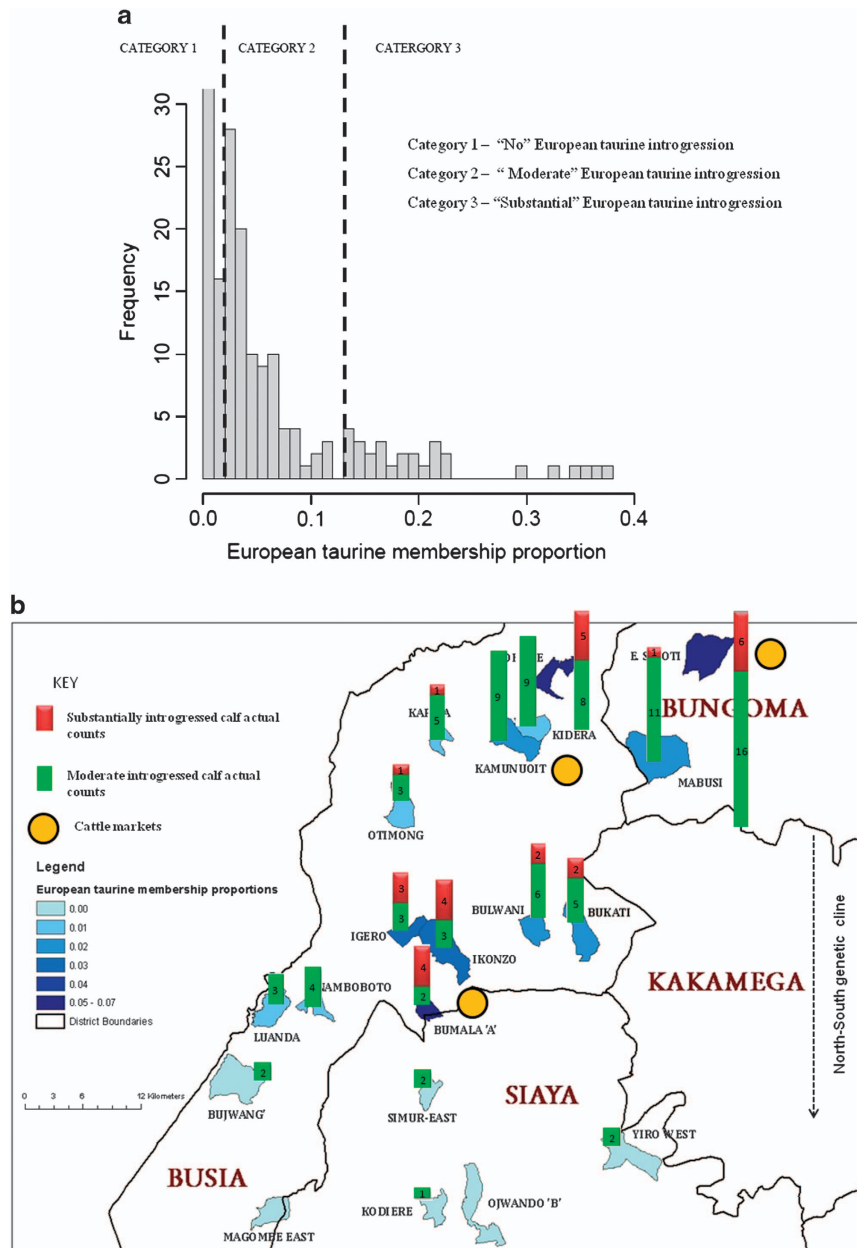


Figure 3 (a) Histogram representing the frequency of calves with varying levels of European ancestry. Three categories of European taurine introgression of data set 1 were defined using the Ward algorithm. (b) Geospatial distribution of the substantial European taurine ($\geq 12.5\%$) and moderate European taurine ($1.56\% < X < 12.5\%$) categories. It indicates a genetic cline and two hotspots of substantial European introgression ($\geq 12.5\%$) found within close proximity of animal markets.

MacHugh *et al.*, 1997). Worth mentioning here, is the detected presence so far of only taurine cattle mitochondrial DNA on the African continent even in populations phenotypically classified as zebu (Gifford-Gonzalez and Hanotte, 2011), is an observation compatible with the pattern of male-mediated zebu introgression into taurine animals or with a selection pressure in favor of taurine mitochondrial haplotypes.

We also assessed if the global genome admixture of the EASZ could be further partitioned to finer detail. In other words, whether the AT and Asian zebu inferred genetic ancestry in EASZ at $K=3$ may be further separated into distinct genome components that reflect in-depth details revealing the history of the breed. Previous studies

indicate that through the Horn of Africa, the African continent likely witnessed two waves of zebu introductions and migrations (Hanotte *et al.*, 2002), which may have distinctively imprinted the genome of the EASZ. The presence of a unique shared genetic background at $K=4$ (Figure 1) between EASZ and Sheko, absent in both the Nelore and N'dama breeds, is compatible with a two wave zebu introgression pattern. It would be tempting to claim that this additional component may represent the first phase intermediate zebu–taurine hybrid, so-called sanga cattle (see Rege, 1999; Rege and Tawah, 1999), but in absence of an appropriate reference population such interpretation remains hypothetical. Indeed, the presence of a ‘unique’ East African genetic background will also be compatible with an indigenous

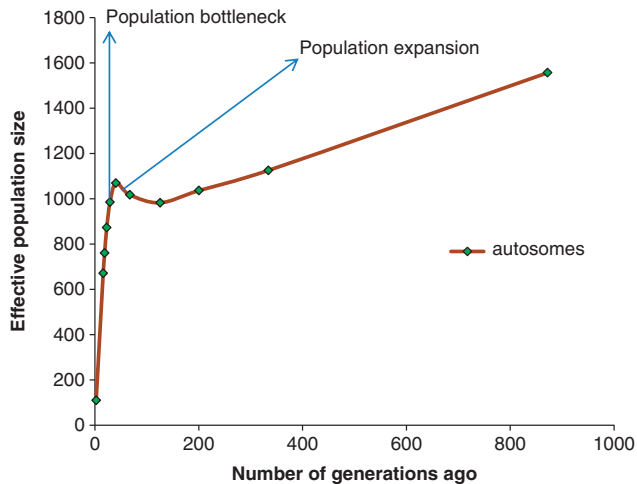


Figure 4 General trend of the average effective population size variation against past generations.

African ancestry (for example, from African wild aurochs introgression). The analysis of more African cattle populations, representative of a broader geographic area from the continent, as well as additional reference breeds drawn outside of the continent may further clarify this issue. Until then, we should favor the clearly interpretable three clusters model previously supported by several studies (for example, Hanotte *et al.*, 2002).

As mentioned above, the level of European taurine ancestry in EASZ was unevenly distributed among calves, but relatively similar among chromosomes (Figure 1 and Supplementary Figure S4a). There are several possible European taurine genetic sources in the studied area. These include concluded and ongoing dairy breed improvement programs that used/use exotic animals and semen. Also, animal markets in the studied area are stocked with crossbred animals. Although the overall average European ancestry was only 2% for the entire studied population, the proportions varied across the study area. The animals with substantial European ancestry ($\geq 12.5\%$), which is compatible with a European taurine introgression event three or less generations ago, were found close to major livestock markets (Figure 3b). Interestingly, the average proportions of European taurine (ranging between 4 and 7%) were higher in northern sub-locations such as Busia and Bungoma that participated actively in initiatives to develop the local dairy production systems (Baltenweck *et al.*, 2005). In the southern sub-locations found within Siaya district, little or no European taurine introgression was observed ($\leq 1.25\%$; Figure 3b). We purposely avoided the sampling of first-generation crossbred animals during the study and therefore our data is not representative of the current impact of dairy sector development activities on the indigenous EASZ population. Even so, our study still highlights the trend of European taurine introgression into the EASZ population. An interesting related question is whether or not such exotic introgression is under selection? Commercial cattle originating from the temperate environment are known to be poorly adapted to most of the tropical agroecosystems (for example, Vordermeier *et al.*, 2012) and selection against exotic introgression will be expected. Our recent analysis of the same EASZ animals suggests that it is indeed the case with increased vulnerability to infectious diseases for the EASZ introgressed animals (Murray *et al.*, 2013b).

Excluding introgressed animals with exotic taurine, nonsignificant genetic differentiation was observed between sub-locations

(Supplementary Figure S6). Low or absence of differentiation implies frequent exchanges and/or movements of animals across a geographical area (Tapio *et al.*, 2010; Dumasy *et al.*, 2012). This is compatible with movement of livestock, including breeding bulls across the studied area and active animal trade between sub-locations. It may also explain why we do observe in two sub-locations (Bukati and Luanda) a few calves with different indigenous ancestral proportion. The parentage from these calves could have included other indigenous breed(s) present near the studied area.

EASZ are believed to comprise of several sub-populations according to the farmer communities rearing them. These include the Kavirondo zebu reared by the Luo and Luhya communities, and the Teso zebu reared by the Teso community (DAGRIS, 2007). Berthouly *et al.* (2009) have shown that, in Vietnamese goat, genetic differentiation is greatly influenced by farmers' ethnicity and husbandry practices. There is no apparent evidence that this may be the case in our study population.

Decrease in LD as a function of distance between markers has been reported in several cattle population or breeds (for example, de Roos *et al.*, 2008; Gautier *et al.*, 2009; Flury *et al.*, 2010). The observed intermediate LD in EASZ, $0.05 < r^2 < 0.2$ (Supplementary Figure S7), reaches approximately as far in the genome as in other cattle populations (de Roos *et al.*, 2008; Gautier *et al.*, 2009; Flury *et al.*, 2010), but both the minimum and maximum r^2 were lower than commonly observed. The background LD (the minimum value to which the mean LD asymptotes with increasing distance between markers) is approximately a quarter of those generally observed in cattle populations (de Roos *et al.*, 2008; Gautier *et al.*, 2009; Flury *et al.*, 2010; Supplementary Figure S7). It is similar to the expected value for a heavily stratified population (Gautier *et al.*, 2007). This may suggest that our population has been subdivided in the past, although today we do not observe genetic differentiation across the studied area.

There are several algorithms for inferring population size from LD and it remains unclear which method is the most appropriate one (for example, Corbin *et al.*, 2012). However, cattle LD estimates indicate that all these populations, which include Asian zebu, AT and European taurine breeds, have been shrinking over time (Gautier *et al.*, 2007; de Roos *et al.*, 2008; The Bovine HapMap Consortium, 2009; Flury *et al.*, 2010). The trend has been associated with domestication events, artificial selection for economic traits and breed formation (The Bovine HapMap Consortium, 2009). We do observe a similar declining trend in population size for the EASZ (Figure 4). However, in contrast to other cattle populations, an increase in effective population size is observed approximately 126 generations ago.

Historical cattle generation lengths, prior the intensive breeding system, are expected to be longer than 4 years (Kidd and Cavalli-Sforza, 1974). We assume a 6 years generation length in the EASZ as estimated in Red Sindhi cattle on the Indian subcontinent (Mahadevan, 1955). It is slightly shorter than 6.72 years estimated by Alim (1960) for Kenana zebu cattle in Sudan. With this assumption, the EASZ expansion would have begun approximately 126 generations or ~ 750 years ago, which is the time when zebu cattle supposedly became common in the Rift Valley (Payne, 1970). We then observe a drastic decrease in effective population size starting around 40 generation or 240 years ago up to present time (Figure 4). During this time span (~ 240 –750 years ago), three exceptionally favorable climatic periods with relatively shorter dry spells were experienced in Eastern Africa (Verschuren *et al.*, 2000). The cattle population seemed to thrive during these favorable seasons but

drastically shrunk during the subsequent Lapanarat–Mahlatule drought that was characterized by a sequel of severe droughts and political upheavals (Verschuren *et al.*, 2000). It is also well documented that East African cattle were decimated following the rinderpest epidemic at the end of the nineteenth century (Blench, 1993; Paynes and Hodges, 1997). Our data suggests that the decline in cattle population in the region had already started before the disease outbreaks.

The BovineSNP50 v. 1 beadchip used in this study allowed us to estimate the ancestry proportions of the three main cattle lineages within the EASZ. The largest proportion of the markers included on this beadchip was selected for informativeness in European taurine breeds (Matukumalli *et al.*, 2009), and not surprisingly, a small number of these markers are sufficient for the estimation of the European ancestral proportion (Supplementary Figure S8a). This ascertainment bias makes the SNP chip particularly suitable for the detection of European taurine introgression in African native cattle populations. In contrast, more markers were required for accurate estimation of the zebu (Supplementary Figure S8b) and more particularly the AT ancestries (Supplementary Figure S8c). Owing to this technical shortcoming, the taurine ancestral estimates of our study population may not be fully conclusive. The current availability of a BovineHD genotyping bead chip (Illumina) offers an opportunity to clarify this issue. In addition, we show here that the neighboring markers in the BovineSNP50 v. 1 beadchip are not typically in strong LD, and the chip is informative only for approximately half of the EASZ genome, making the chip an incomplete tool for genome-wide association study, detection of signatures of selection and genomic selection.

The history of East African cattle is complex. It was closely intermingled with the history of local human communities (Hanotte *et al.*, 2002). Undoubtedly, the need to adapt to the harsh tropical environments of the area must have shaped the present day East African cattle genomes. This article presents valuable insights toward better understanding the genetic landscape, genetic affinities and demographic history of an African indigenous cattle breed as well as highlights the urgent need to implement crucial management strategies pertinent in the population's sustenance. The expected 'Livestock Revolution' (Delgado *et al.*, 1999) describes a growing demand for animal products especially in developing countries. It presents a unique opportunity to harness the African indigenous livestock productivity potential by re-defining current breeding strategies aimed at obtaining both productive and resilient animals, while concurrently preserving the rich genetic variability of these populations. In-depth characterization of the genome of these indigenous African breeds is an essential step toward achieving these ultimate goals.

DATA ARCHIVING

Data available from the Dryad Digital Repository: doi:10.5061/dryad.bc598.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We extend our sincere gratitude to the Wellcome Trust (grant reference 07995) for financially supporting this project. USDA-ARS bovine functional laboratory and Geneseek veterinary diagnostics provided invaluable technical assistance through the genotyping of the samples. This project was also partially supported by USDA Agricultural Research Service project 1265-31000-104-00D. Mention of trade names or commercial products in this article

is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. We also thank Professor Steve Kemp (University of Liverpool and ILRI) for providing laboratory space and technical support through the help of Wasike Selieli, Moses Ogugo and John Wambugu. We also wish to acknowledge the grass root farmers of Western Kenya who participated fully and made this project a success. Last but not least, we thank the referees for their constructive comments and suggestions during the reviewing process.

- Alim KA (1960). Reproductive rates and milk yield of Kenana cattle in Sudan. *J Agricultural Sci* **55**: pp 183–188.
- Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**: 1294–1296.
- Baltenweck I, Staal S, Njoroge L, Wanyoike F, Ochungo P, Kariuki E (2005). *Targeting Pro-Poor Investment in the Kenyan Dairy Sub-Sector. Report prepared by the International Livestock Research Institute (ILRI) for IFAD and the Government of Kenya*. ILRI: Nairobi, Kenya.
- Barendse W, Harrison BE, Bunch RJ, Thomas MB, Turner LB (2009). Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. *BMC Genomics* **10**: 178.
- Bebe BO, Udo HJM, Rowlands GJ, Thorpe W (2003). Smallholder dairy systems in the Kenyan highlands: breed preferences and breeding practices. *Livest Prod Sci* **82**: 117–127.
- Berthouly C, Do Ngoc D, Thévenon S, Bouchel D, Van TN, Danes C *et al.* (2009). How does farmer connectivity influence livestock genetic structure? A case-study in a Vietnamese goat population. *Mol Ecol* **18**: 3980–3991.
- Blench RM, MacDonald KC (2000). *The Origins and Development of African Livestock: Archeology, Genetics, Linguistics and Ethnography*. Blench RM, MacDonald KC (eds). Routledge: New York, pp 566.
- Blench R (1993). Ethnographic and linguistic evidence for the prehistory of African ruminant, livestock, horses and ponies. In: Shaw T, Sinclair P, Andah B, Okpoko A (eds). *The Archeology of Africa*, Routledge: London, pp 71–103.
- Bonfiglio S, Ginja C, De Gaetano A, Achilli A, Olivieri A, Colli L *et al.* (2012). Origin and Spread of *Bos taurus*: new clues from mitochondrial genomes belonging to haplogroup T1. *PLoS ONE* **7**: e38601.
- Bonsma JC (1973). Crossbreeding for adaptability. In: Koger M, Cunha TJ, Warnick AC (eds). *Crossbreeding Beef Cattle, Series 2*, University of Florida Press: Gainesville, FL, USA.
- Bradley DG, MacHugh DE, Cunningham P, Loftus RF (1996). Mitochondrial diversity and origin of African and European cattle. *Proc Natl Acad Sci USA* **93**: 5131–5135.
- Carvalho FA, Lammoglia MA, Simoes MJ, Randel RD (1995). Breed affects thermoregulation and epithelial morphology in imported and native cattle subjected to heat stress. *J Anim Sci* **73**: 3570–3573.
- Corbin LJ, Liu AYH, Bishop SC, Woolliams JA (2012). Estimation of historical effective population size using linkage disequilibria with marker data. *J Anim Breed Genet* **129**: 257–270.
- DAGRIS (2007). *Domestic Animal Genetic Resources Information System (DAGRIS)*. Kemp S, Mamo Y, Asrat B, Dessie T (eds). International Livestock Research Institute: Addis Ababa, Ethiopia. <http://dagris.ilri.cgiar.org>.
- Delgado C, Rosegrant M, Steinfeld H, Ehui S, Courbois C (1999). *Livestock to 2020: The Next Food Revolution. IFPRI Food, Agriculture, and the Environment Discussion Paper 28*. IFPRI: Washington, DC, USA.
- de Roos APW, Hayes BJ, Spelman R, Goddard ME (2008). Linkage disequilibrium and persistence of phase in Holstein Friesian, Jersey and Angus cattle. *Genetics* **179**: 1503–1512.
- Dray S, Dufour AB (2007). The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* **22**: 1–20.
- Dumasy JF, Daniaux C, Donnay I, Baret PV (2012). Genetic diversity and networks of exchange: a combined approach to assess intra-breed diversity. *Genet Sel Evol* **44**: 17.
- Epstein HE (1971). *The Origin of the Domestic Animals of Africa*, vol. 1. Africana Publishing Corporation: New York, USA, pp 573.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620.
- Falush D, Stephens M, Pritchard JK (2003). Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.
- Flury C, Tapio M, Sonstegard T, Drögemüller C, Leeb T, Simianer H, Hanotte O, Rieder S (2010). Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *J Anim Breed Genet* **127**: 339–347.
- Gaughan JB, Mader TL, Holt SM, Josey MJ, Rowan KJ (1999). Heat tolerance of Boran and Tuli crossbred steers. *J Anim Sci* **77**: 2398–2405.
- Gautier M, Faraut T, Moazami-Goudarzi K, Navratil V, Foglio M, Grohs C *et al.* (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* **177**: 1059–1070.
- Gautier M, Flori L, Riebler A, Jaffrezic F, Laloe D, Gut I *et al.* (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* **10**: 550.
- Gautier M, Laloe D, Moazami-Goudarzi K (2010). Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS ONE* **5**: e13038.

- Gautier M, Naves M (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol Ecol* **20**: 3128–3143.
- Gifford-Gonzalez D, Hanotte O (2011). Domesticating animals in Africa: implications of genetic and archaeological findings. *J World Prehistory* **24**: 1–23.
- Goudet J (2005). *Hierfstat*, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes* **5**: 184–186.
- Haldane JBS (1919). The recombination of linkage values and the calculation of distance between linked factors. *J Genet* **8**: 299–309.
- Haile A, Workneh A, Noah K, Tadello D, Azage T (2011). Breeding strategy to improve Ethiopian Boran cattle for meat and milk production. *IPMS (Improving Productivity and Market Success) of Ethiopian Farmers Project Working Paper 26*. ILRI: Nairobi, Kenya.
- Hammond AC, Olson TA, Chase Jr CC, Bowers EJ, Randel RD, Murphy CN *et al.* (1996). Heat tolerance in two tropically adapted *Bos taurus* breeds, Senepol and Romosinuano, compared with Brahman, Angus, and Hereford cattle in Florida. *J Anim Sci* **74**: 295–303.
- Hanotte O, Tawah CL, Bradley DG, Okomo M, Verjee Y, Ochieng J *et al.* (2000). Geographical distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds. *Mol Ecol* **9**: 387–396.
- Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JEO (2002). African pastoralism: genetic imprints of origins and migrations. *Science* **296**: 336–339.
- Hanotte O, Ronin Y, Agaba M, Nilsson P, Gelhaus A, Horstmann R *et al.* (2003). Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc Natl Acad Sci USA* **97**: 7443–7448.
- Hansen PJ (2004). Physiological and cellular adaptations of zebu cattle to thermal stress. *Anim Reprod Sci* **83**: 349–360.
- Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res* **13**: 635–643.
- Hill WG, Robertson A (1968). Linkage disequilibrium in finite populations. *Theor Appl Genet* **38**: 226–231.
- Jakobsson M, Rosenberg NA (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.
- Jombart T (2008). ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.
- Kidd KK, Cavalli-Sforza LL (1974). The role of genetic drift in the differentiation of Icelandic and Norwegian cattle. *Evolution* **28**: 381–395.
- Latif AA, Punyua DK, Nokoe S, Capstick PB (1991a). Tick infestations on Zebu cattle in western Kenya: individual host variation. *J Med Entomol* **28**: 114–121.
- Latif AA, Punyua DK, Nokoe S, Capstick PB (1991b). Tick infestations on Zebu cattle in western Kenya: quantitative assessments of host resistance. *J Med Entomol* **28**: 122–126.
- Lawrence JA, Musisi FL, Mfitilozde MW, Tjorneho JK, Whiteland AP, Kafuwa PT *et al.* (1996). Integrated tick and tick-borne disease control trials in crossbred dairy cattle in Malawi. *Trop Anim Health Prod* **28**: 280–288.
- Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P (1994). Evidence for two independent domestications of cattle. *Proc Natl Acad Sci USA* **91**: 2757–2761.
- Mahadevan P (1955). Population and production characteristics of Red Sindhi cattle in Ceylon. *J Dairy Sci* **38**: 1231–1241.
- MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG (1997). Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146**: 1071–1086.
- Mattioli RC, Bah M, Faye J, Kora S, Cassama M (1993). A comparison of field tick infestation on N'dama, Zebu and N'dama x Zebu crossed cattle. *Vet Parasitol* **47**: 139–148.
- Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP *et al.* (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* **4**: e5350.
- McTavish EJ, Decker JE, Schnabel RD, Taylor JF, Hillis DM (2013). New world cattle show ancestry from multiple independent domestication events. *Proc Natl Acad Sci USA* **110**: E1398–E1406.
- Melka MG, Schenkel FS (2012). Analysis of genetic diversity of Brown Swiss, Jersey and Holstein populations using genome-wide single nucleotide polymorphism markers. *BMC Res Notes* **5**: 161.
- Murray GGR, Woolhouse MEJ, Tapio M, Mbole-Kariuki MN, Sonstegard TS, Thumbi SA *et al.* (2013a). Data from: genetic susceptibility to infectious disease in East African Shorthorn Zebu: a genome-wide analysis of the effect of heterozygosity and exotic introgression. Dryad Digital Repository 10.5061/dryad.bc598.
- Murray GGR, Woolhouse MEJ, Tapio M, Mbole-Kariuki MN, Sonstegard TS, Thumbi SA *et al.* (2013b). Genetic susceptibility to infectious disease in East African Shorthorn Zebu: a genome-wide analysis of the effect of heterozygosity and exotic introgression. *BMC Evol Biol* **13**: 246. <http://dx.doi.org/10.1186/1471-2148-13-246>.
- Pant SD, Schenkel FS, Verschoor CP, You QM, Kelton DF, Moore SS *et al.* (2010). A principal component regression based genome wide analysis approach reveals the presence of a novel QTL on BTA7 for MAP resistance in Holstein cattle. *Genomics* **95**: 176–182.
- Payne WJA (1970). Cattle production in the tropics. In: *Breeds and Breeding*, vol. 1. Longman: London, UK.
- Paynes WJA, Hodges J (1997). *Tropical Cattle: Origins, Breeds and Breeding policies*. Blackwell Science: Oxford.
- Pritchard KJ, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Pritchard KJ, Wen X, Falush D (2007). Documentation for structure software; Version 2.2. <http://pritch.bsd.uchicago.edu/software/structure22/>.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.
- Rege JEO (1999). The state of African cattle genetic resources I. Classification framework and identification of threatened and extinct breeds. *Animal Genetic Resources Information* **25**: 1–25.
- Rege JEO, Tawah CL (1999). The state of African cattle genetic resources II. Geographical distribution, characteristics and uses of present-day breeds and strains. *Animal Genetic Resources Information* **26**: 1–25.
- Rege JEO, Kahi A, Okomo-Adhiambo M, Mwacharo J, Hanotte O (2001). *Zebu Cattle of Kenya: Uses, Performance, Farmer Preferences, Measures of Genetic Diversity and Options for Improved Use*. *Animal Genetic Resources Research 1*. ILRI (International Livestock Research Institute): Nairobi, Kenya, pp 103.
- Rousset F (1997). Genetic differentiation and estimation of gene flow for F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- Settles M, Zanella R, McKay SD, Schnabel RD, Taylor JF, Whitlock R *et al.* (2009). A whole genome association analysis identifies loci associated with *Mycobacterium avium* subsp. paratuberculosis infection status in US Holstein cattle. *Anim Genet* **40**: 655–662.
- Tapio M, Ozerov M, Tapio I, Toro MA, Marzanov N, Ćinkulov M *et al.* (2010). Microsatellite-based genetic diversity and population structure of domestic sheep in northern Eurasia. *BMC Genet* **11**: 76.
- The Bovine HapMap Consortium (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**: 528–532.
- Turner JW (1980). Genetic and biological aspects of Zebu adaptability. *J Anim Sci* **50**: 1201–1205.
- Van Tassel CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT *et al.* (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods* **5**: 247–252.
- Verschuren D, Laird KR, Cumming BF (2000). Rainfall and drought in equatorial east Africa during the past 1,100 years. *Nature* **403**: 410–414.
- Vordermeier M, Ameni G, Berg S, Bishop R, Robertson BD, Aseffa A *et al.* (2012). The influences of cattle breed on susceptibility to bovine tuberculosis in Ethiopia. *Comp Immunol Microbiol Infect Dis* **35**: 227–232.
- Wambura PN, Gwakisa PS, Silayo RS, Rugaimukamu EA (1998). Breed-associated resistance to tick infestation in *Bos indicus* and their crosses with *Bos taurus*. *Vet Parasitol* **77**: 63–70.
- Ward JH (1963). Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* **58**: 236–244.
- Weir BS, Hill WG (1980). Effect of mating structure on variation in linkage disequilibrium. *Genetics* **95**: 477–488.
- Weir A, Nottley M, Katui-Katua M (2009). *Finnish Aid in Western Kenya. Impact and Lessons Learned. Evaluation report 2009:5*. Ministry for Foreign Affairs of Finland. Hakapaino Oy: Helsinki, pp 107.
- Wiggins GR, Sonstegard TS, VanRaden PM, Matukumalli LK, Schnabel RD, Taylor JF *et al.* (2008). Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *J Dairy Sci* **92**: 3431–3436.
- Wiggins GR, VanRaden PM, Bacheller LR, Tooker ME, Hutchison JL, Cooper TA *et al.* (2010). Selection and management of DNA markers for use in genomic evaluation. *J Dairy Sci* **93**: 2287–2292.
- World Geodetic System (WGS84) (1984). <http://www.ngi.gov.za/index.php/technical-information/geodesy-and-gps/world-geodetic-system-1984-wgs84-and-the-international-terrestrial-reference-frame-itrf>.



This work is licensed under a Creative Commons Attribution 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)