

# Revisiting demographic processes in cattle with genome-wide population genetic analysis

Pablo Orozco-terWengel<sup>1\*</sup>, Mario Barbato<sup>1</sup>, Ezequiel Nicolazzi<sup>2</sup>, Filippo Biscarini<sup>2</sup>, Marco Milanesi<sup>3</sup>, Wyn Davies<sup>4</sup>, Don Williams<sup>5</sup>, Alessandra Stella<sup>2</sup>, Paolo Ajmone-Marsan<sup>3</sup> and Michael W. Bruford<sup>1</sup>

<sup>1</sup> School of Biosciences, Cardiff University, Cardiff, UK, <sup>2</sup> Parco Tecnologico Padano, Lodi, Italy, <sup>3</sup> Faculty of Agriculture, Università Cattolica del Sacro Cuore, Piacenza, Italy, <sup>4</sup> Dinefwr, National Trust, Llandeilo, UK, <sup>5</sup> Independent Researcher, Swansea, UK

## OPEN ACCESS

### Edited by:

Peter Dovc,  
University of Ljubljana, Slovenia

### Reviewed by:

Ino Curik,  
University of Zagreb, Croatia  
Linyang Xu,  
University of Maryland, USA

### \*Correspondence:

Pablo Orozco-terWengel,  
School of Biosciences, Cardiff  
University, Sir Martin Evans Building,  
Museum Avenue, Cardiff CF10 3AX,  
UK  
orozco-terwengel@cardiff.ac.uk

### Specialty section:

This article was submitted to  
Livestock Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 12 February 2015

**Paper pending published:**  
26 March 2015

**Accepted:** 12 May 2015

**Published:** 02 June 2015

### Citation:

Orozco-terWengel P, Barbato M, Nicolazzi E, Biscarini F, Milanesi M, Davies W, Williams D, Stella A, Ajmone-Marsan P and Bruford MW (2015) Revisiting demographic processes in cattle with genome-wide population genetic analysis. *Front. Genet.* 6:191. doi: 10.3389/fgene.2015.00191

The domestication of the aurochs took place approximately 10,000 years ago giving rise to the two main types of domestic cattle known today, taurine (*Bos taurus*) domesticated somewhere on or near the Fertile Crescent, and indicine (*Bos indicus*) domesticated in the Indus Valley. However, although cattle have historically played a prominent role in human society the exact origin of many extant breeds is not well known. Here we used a combination of medium and high-density Illumina Bovine SNP arrays (i.e., ~54,000 and ~770,000 SNPs, respectively), genotyped for over 1300 animals representing 56 cattle breeds, to describe the relationships among major European cattle breeds and detect patterns of admixture among them. Our results suggest modern cross-breeding and ancient hybridisation events have both played an important role, including with animals of indicine origin. We use these data to identify signatures of selection reflecting both domestication (hypothesized to produce a common signature across breeds) and local adaptation (predicted to exhibit a signature of selection unique to a single breed or group of related breeds with a common history) to uncover additional demographic complexity of modern European cattle.

**Keywords:** cattle, *Bos taurus*, *Bos indicus*, phylogeography, adaptation, SNP array

## Introduction

Modern cattle comprise two species, *Bos taurus* and *Bos indicus*, both derived from the extinct aurochs (*B. primigenius*). The divergence between the two species has been dated ~250,000 years ago on the basis of mitochondrial DNA haplotypes (Bradley et al., 1996). The two species are interfertile and a number of hybrid breeds have been established (e.g., the American Beefmaster that was established from a Shorthorn × Hereford × Brahman cross). The details of cattle domestication have been a contentious, and arguments in favor of two or three domestication events have been put forward. The hypothesis arguing for two domestication sites suggests that taurine cattle were domesticated in the Fertile Crescent, likely from *B. primigenius*, while indicine cattle were domesticated in the Indus valley, likely from *B. primigenius namadicus* (Loftus et al., 1994; Troy et al., 2001; Lewis et al., 2011). The hypothesis arguing for three domestication sites adds a Northeast African center, that gave rise to the African taurine breeds deriving from *B. primigenius opisthonomus* (Wendorf and Schild, 1994; Payne and Hodges, 1997). Mitochondrial DNA evidence suggests that two main lineages were domesticated, one now typical of taurine

breeds, and the other now typical of indicines (Loftus et al., 1994; Bradley et al., 1996). The domestication process is thought to have originated sometime between 8000 and 10,000 years ago (Ajmone-Marsan et al., 2010; Gautier et al., 2010; Lewis et al., 2011). Several potential routes have been hypothesized to explain the dispersal of taurine cattle from the Fertile crescent, mediated by Neolithic farmers, including alternative routes into Europe, occurring on at least two occasions (a Danubian route and a Mediterranean route Payne and Hodges, 1997; Pellecchia et al., 2007; Ajmone-Marsan et al., 2010; Taberlet et al., 2011). As aurochs only became extinct recently (Machugh et al., 1998; Beja-Pereira et al., 2006), it is possible that both taurine and indicine cattle crossed with their wild ancestor during their expansion out of the domestication centers, although the evidence supporting it is much debated (Troy et al., 2001; Gotherstrom et al., 2005; Beja-Pereira et al., 2006; Bollongino et al., 2008; Perez-Pardal et al., 2010; Edwards et al., 2011).

Several studies have addressed the population structure and phylogenetic relationships between taurine and indicine cattle breeds. Initially, studies were based on allozymes, mitochondrial DNA sequences, and microsatellites (Loftus et al., 1994; Medjugorac et al., 1994; Bradley et al., 1996; Machugh et al., 1998; Kantanen et al., 2000; Hanotte et al., 2002; Gautier et al., 2007) describing patterns of genetic variation in cattle, e.g., the two main lineages and differentiation between taurine and indicine breeds, albeit largely focusing on taurines. Nevertheless, such studies identified a clear differentiation between breeds from the two species, and between African and European taurines (Gautier et al., 2007). Interestingly, this handful of markers separated the African taurine breeds (e.g., Lagune, N'Dama and Somba) living in tsetse fly endemic regions from those that did not co-occur with these flies (e.g., Kuri and Borgou), suggesting that demographic signals among these breeds are strong and possibly related to properties of the environment they occupy. Recently, surveys of genome-wide genetic variation in many breeds have been carried out using SNP arrays (Bovine Hapmap et al., 2009; Mctavish et al., 2013; Decker et al., 2014). These more comprehensive analyses included tens of thousands of SNPs distributed across all chromosomes genotyped in several breeds from around the world. With these data, the differentiation between taurine and indicine breeds has been measured at approximately 10%, with approximately 5% for the divergence between African and non-African taurines (Gautier et al., 2010; Mctavish et al., 2013; Decker et al., 2014). Analyses using this type of large-scale data further identified a split between African longhorn taurines (N'Dama) and shorthorn taurines (e.g., Somba and Lagune—Gautier et al., 2010). Such analyses have also proved powerful enough to identify the mixed indicine/taurine genetic component of recently established hybrid breeds such as Santa Gertrudis and Beefmaster (Gautier et al., 2010; Mctavish et al., 2013).

Cattle are the most common large livestock species in the world; the global population size of which is approximately 1400 million animals (Feliuss et al., 2011; Taberlet et al., 2011). Among these, 1311 breeds have been recognized, 209 of which are now extinct (Risckowsky and Pilling, 2007). Since domestication, cattle have spread around the world, with farmers selecting

animals with desirable characteristics and establishing many breeds. This process, while not passive, was until recently slow and resulted in animals becoming adapted to the local conditions, e.g., feed types, local weather, and diseases (Wood, 1973; Cobb, 2006; Russell, 2007). The last two centuries saw a dramatic acceleration in the process of artificial/directional selection, largely as a consequence of the development of the breed concept, artificial insemination, and the development of statistical approaches used to estimate breeding values used to choose sires and dams for breeding purpose. This has limited the intercrossing between animals of divergent genetic background in order to increase or maintain breed purity, especially in dairy cattle [in beef cattle, conversely, cross-breeding is often used to exploit also non-additive genetic variation i.e., heterosis (Simm, 1998); still, pure lines for cross-breeding need to be selected and maintained]. Additionally, as a consequence of artificial insemination and marketing policies, few “champion” males with characteristics of interest have been used extensively to fertilize large proportions of females. The above developments have led to: (i) intense within-breed (or line) selection; (ii) importation/exportation of elite breeding stock to different environments/countries (migration); (iii) strong founder effect of “champion/top ranked” males both in breed formation or in already established breeds whose effective population size has thus been shrunken; (iv) larger genetic drift as a consequence of breed isolation. Generally, by trying to keep breeds pure, gene-flow among them has been reduced or even halted (unless done to improve breeds by upgrading Feliuss et al., 2011), and by using relatively few males to fertilize many females, inbreeding is increased (Taberlet et al., 2011). These approaches are used to manipulate the properties of a breed at the expense of the decrease in the genetic variability (Bulmer, 1971) which may be accompanied by a decrease in genetic health of the breed (e.g., reduction in fertility in Holstein cattle via the increase in frequency of deleterious variants through the process of inbreeding) (Pryce et al., 1997, 2004; Biffani et al., 2005). Overall, this change in farming practices resulted in a change from limited selection throughout much of the domestic history of cattle (~9800 years) to relatively stronger selection in the last ~200 years.

The relatively rapid distribution of cattle around the world over the last ~10,000 years suggests that populations have been potentially exposed to selective pressures deriving from the environmental variables that they had not been exposed previously, e.g., diseases occurring outside of the domestication centers (Hanotte et al., 2002; Feliuss et al., 2014; Xu et al., 2014). Consequently, while some cattle breeds represent recent introductions to new geographical areas such as the New World or Oceania, other breeds represent long established populations outside of the domestication center. The genetic makeup of such populations is a valuable source to understand the processes driving local adaptation, as the former represent genetic pools currently being shaped by local selective pressures (i.e., may serve as examples of selection in action) (Flori et al., 2012), while the latter represent established populations where the extant genetic variation reflects the outcome of local adaptation (Feliuss et al., 2014). In particular, long established local breeds are likely to represent the reservoirs of important genetic variation

with adaptive potential that needs to be characterized (Taberlet et al., 2011; Felius et al., 2014) before it possibly disappears as a consequence of the widespread use of a handful of industrial breeds among others (Fao, 2007; Thornton, 2010; Herrero and Thornton, 2013; Perry et al., 2013).

Here, we collated a large-scale genome-wide dataset of SNP polymorphism in cattle breeds from around the world representing published taurine and indicine breeds, and new data for four European taurine breeds, all genotyped for a common set of ~35,000 SNPs in over 1300 animals. We used these data to characterize the distribution of genetic variation among cattle breeds around the world, and the genealogical relationships between these breeds. The dataset described here represents multiple breeds from various continents (e.g., multiple European and African taurines, as well as Asian and African indicines) providing a unique experimental set up that allows searching for signatures of selection that are consistent across multiple breeds from different continents. This approach enables to separate signatures of selection that may be breed specific from selection signals common for several breeds occurring in the same geographic region. We extend the published dataset with new SNP data to a case study where we compare white cattle breeds of conservation concern in the UK (White Park and Chillingham) with a candidate Roman ancestor, the Chianina, to help resolve the commonly held belief of a Roman origin of the Welsh White Park cattle (Felius et al., 2011).

## Materials and Methods

### SNP Array Data

SNP array data was obtained for a total of 1346 animals representing 46 cattle breeds, and four species related to cattle (*B. gaurus*—Gaur, *B. javanicus*—Banteng, *B. gruniensis*—Yak, and *Bison bison*—Bison) used as outgroups (Table 1). The data comprises a combination of previously published datasets and new data collected for this study. The published data represents breeds published by Gautier et al. (2009); Gautier et al. (2010) and the Bovine HapMap project (Bovine Hapmap et al., 2009) genotyped with Illumina's BovineSNP50 v.1 and v.2 chip assay, respectively (Bovine Hapmap et al., 2009) (Table 1). Four new breeds were genotyped with the BovineHD Beach Chip for this project, namely, Welsh White Park (Dinefwr), Chianina, Romagnola and Chillingham. All SNP coordinates were converted to UMD3.1 bovine assembly (RefSeq: GCF\_000003055.5). The datasets were merged using PLINK v1.7 (Purcell et al., 2007). The merged dataset was filtered to keep only individuals with 95% of their SNPs called and SNPs with 95% call rate across all samples. The SNPs and individuals that did not pass these filters were discarded. The total SNPs left in the dataset was 36,503 (data is available upon request from the authors). The data for each breed was phased with fastPHASE v1.4 (Scheet and Stephens, 2006).

### Population Divergence

An analysis of population structure was carried out using the software Admixture v1.22 (Alexander et al., 2009), which uses a model-based estimation of individual ancestry for a range

of prior values of  $K$  defined by the user. A cross-validation approach is used to determine the most likely number of populations ( $K$ ) in the data. For each tested value of  $K$ , Admixture estimates the proportion of each individual's genotype deriving from each cluster. The values of  $K$  tested were in the range between 1 and 60 to accommodate potential population structure within breeds. For this analysis we excluded outgroup samples, breeds represented by fewer than 10 individuals, and SNPs with a minimum allele frequency lower than 0.01, linkage disequilibrium higher than 0.1 using a sliding window approach of 50 SNPs and a step size of 10 SNPs (Alexander et al., 2009). A principal components analysis was carried out using flashpca v1.2 (Abraham and Inouye, 2014) with default settings in order to investigate the ordinal relationships among breeds and individuals. Lastly, a NeighbourNet network was constructed using Reynold's distance among breeds using Splitstree v4.13.1 (Huson and Bryant, 2006), and a dendrogram using the same genetic distance and 100 bootstrap replicates to assess the statistical support of breed; clustering was performed using Phylip v3.69 (Felsenstein, 1993).

Within this framework, the Welsh White Park cattle breed was analyzed along the other cattle breeds to identify their relative position within the dendrogram and PCA explaining the similarities between cattle breeds. In particular, the common belief of a Roman origin of the Welsh White Park was addressed by comparing the position of this breed in relation to the Italian breeds, in particular the large Chianina breed, one of the oldest cattle breeds originating from the region of Valdichiana in central Italy, and Romagnola. Additionally, we estimated the relationships between Welsh White Park and other 17 breeds in the dataset using Treemix v1.12 (Pickrell and Pritchard, 2012) in order to determine the historical relationships between these breeds in terms of splits and migration (mixtures) between breeds. The breeds used for this analysis were Welsh White Park, the Italian Chianina, Romagnola (ROM), and Piedmontese, the British breeds Chillingham, Hereford, Angus, Jersey (JER), as well as five other non-African taurines (Holstein (HOL), Brown Swiss, Charolais, Normande, Vosgienne), the four African taurines Baoule, Lagune, Somba and N'Dama (NDA), and the indicine breed Brahman as outgroup. Treemix was run iteratively for values of the migration parameter between 0 and 12. The  $f$  index representing the fraction of the variance in the sample covariance matrix ( $\hat{W}$ ) accounted for by the model covariance matrix ( $W$ ) was used to identify the number of modeled migration events that best fitted the data (Pickrell and Pritchard, 2012).

### Demographic Estimation

Recent demographic history was measured by the trend in effective population size ( $N_e$ ) change over the last 8000 years using the software SNeP v1.0 and default settings (Barbato et al., 2015). Loci with missing data were excluded from this analysis, as well as SNPs with a minimum allele frequency (MAF) smaller than 5% (Sved et al., 2008; Corbin et al., 2012). Linkage disequilibrium (LD) was calculated between each pair of SNPs separated by a minimum distance of 5 kbp and maximum distance of 1 Mbp using Hill and Robertson's  $r^2$

**TABLE 1 | Breed abbreviations and species identification.**

Abb	BN	Tax	D	N	Ho	F
ABO	Abondance	<i>Bos taurus</i>	2	22	0.308	0.089
ANG	Angus	<i>Bos taurus</i>	1	61	0.305	0.098
AUB	Aubrac	<i>Bos taurus</i>	2	22	0.301	0.112
BPN	Bretonne black pied	<i>Bos taurus</i>	2	18	0.317	0.064
BRU	French brown swiss	<i>Bos taurus</i>	2	18	0.288	0.150
BSW	Brown swiss	<i>Bos taurus</i>	1	24	0.280	0.171
CHA	Charolais	<i>Bos taurus</i>	1	20	0.314	0.071
CHL	Charolais	<i>Bos taurus</i>	2	26	0.320	0.054
GAS	Gascon	<i>Bos taurus</i>	2	22	0.305	0.100
GNS	Guernsey	<i>Bos taurus</i>	1	21	0.275	0.188
HFD	Hereford	<i>Bos taurus</i>	1	31	0.304	0.101
HO2	Holstein	<i>Bos taurus</i>	2	31	0.316	0.066
HO1	Holstein	<i>Bos taurus</i>	1	30	0.313	0.077
HOL	Holstein	<i>Bos taurus</i>	2	64	0.319	0.057
JE2	Jersey	<i>Bos taurus</i>	2	28	0.263	0.223
JEF	Jersey	<i>Bos taurus</i>	1	21	0.277	0.180
JER	Jersey	<i>Bos taurus</i>	2	28	0.263	0.223
LMS	Limousin	<i>Bos taurus</i>	1	44	0.309	0.086
MAN	Maine-Anjou (Rouge des Près)	<i>Bos taurus</i>	2	16	0.303	0.105
MAR	Maraichine (Parthenaise)	<i>Bos taurus</i>	2	19	0.318	0.059
MON	Montbeliarde	<i>Bos taurus</i>	2	30	0.299	0.116
NOR	Normande	<i>Bos taurus</i>	2	30	0.307	0.094
NRC	Norwegian red cattle	<i>Bos taurus</i>	1	21	0.317	0.063
OUL	Oulmès Zaer	<i>Bos taurus</i>	2	27	0.288	0.149
PMT	Piedmontese	<i>Bos taurus</i>	1	24	0.321	0.053
PRP	French red pied lowland	<i>Bos taurus</i>	2	22	0.325	0.039
RGU	Red Angus	<i>Bos taurus</i>	1	15	0.305	0.100
RMG	Romagnola	<i>Bos taurus</i>	1	24	0.291	0.141
ROM	Romagnola	<i>Bos taurus</i>	3	13	0.293	0.134
CHI	Chianina	<i>Bos taurus</i>	3	14	0.285	0.158
CIL	Chillingham	<i>Bos taurus</i>	3	16	0.026	0.924
W_P	White park	<i>Bos taurus</i>	3	15	0.245	0.276
SAL	Salers	<i>Bos taurus</i>	2	22	0.285	0.157
TAR	Tarine	<i>Bos taurus</i>	2	18	0.300	0.113
VOS	Vosgienne	<i>Bos taurus</i>	2	20	0.312	0.078
BAO	Baoule	<i>Bos taurus</i>	2	29	0.216	0.362
LAG	Lagune	<i>Bos taurus</i>	2	30	0.183	0.460
NDA	N'Dama	<i>Bos taurus</i>	1	25	0.209	0.381
ND1	N'Dama	<i>Bos taurus</i>	2	14	0.235	0.307
ND2	N'Dama	<i>Bos taurus</i>	2	17	0.237	0.299
ND3	N'Dama	<i>Bos taurus</i>	2	25	0.210	0.381
SOM	Somba	<i>Bos taurus</i>	2	30	0.217	0.359
BRM	Brahman	<i>Bos indicus</i>	1	25	0.190	0.440
GIR	Gir	<i>Bos indicus</i>	1	24	0.160	0.528
NEL	Nelore	<i>Bos indicus</i>	1	21	0.161	0.524
ZBO	Zebu Bororo	<i>Bos indicus</i>	2	23	0.240	0.292
ZFU	Zebu Fulani	<i>Bos indicus</i>	2	30	0.241	0.289
ZMA	Zebu from Madagascar	<i>Bos indicus</i>	2	30	0.194	0.427
BMA	Beefmaster	hybrid	1	24	0.328	0.031
SGT	Santa Gertrudis	hybrid	1	24	0.313	0.075

(Continued)

**TABLE 1 | Continued**

Abb	BN	Tax	D	N	Ho	F
BOR	Borgou	hybrid	2	30	0.263	0.222
KUR	Kuri	hybrid	2	30	0.261	0.229
SHK	Sheko	hybrid	1	20	0.250	0.260
SIM	Simmental	<i>Bos taurus</i>	1	3	NA	NA
GBV	Gelbvieh	<i>Bos taurus</i>	1	3	NA	NA
OBB	North American Bison	<i>Bos Bison</i>	2	4	NA	NA
OBJ	Banteng	<i>Bos javanicus</i>	1	2	NA	NA
OGR	Gaur	<i>Bos gaurus</i>	1	4	NA	NA
OYK	Yak	<i>Bos grunniens</i>	1	2	NA	NA

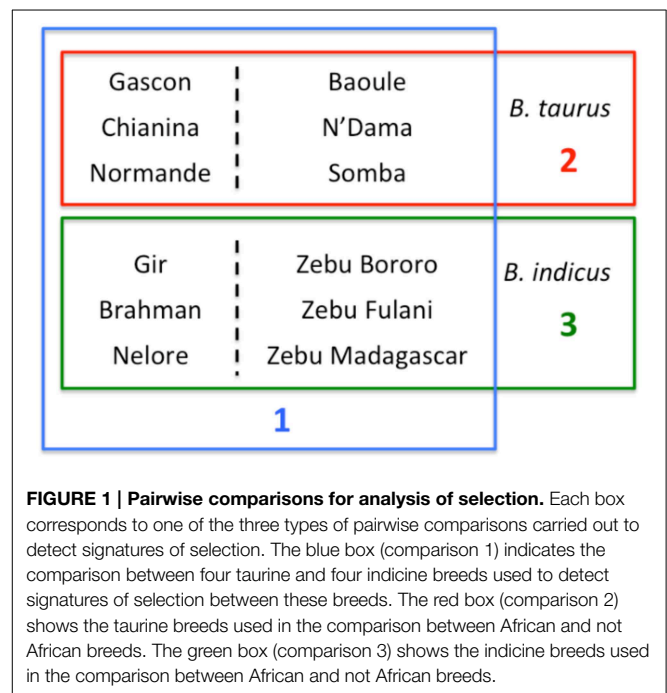
The three letter code used to identify the breed throughout the manuscript is shown (**Abb**), along with the breed name (**BN**) and the taxonomic status indicating the species name or whether the breed is a *B. taurus* x *B. indicus* hybrid (**Tax**), the data set's origin (**D**), namely Bovine Hapmap Consortium (1), Gautier et al., 2010 (2), this study (3), the breed's sample size (**N**), the observed heterozygosity (**Ho**), and the inbreeding coefficient (**F**).

(Hill and Robertson, 1968). LD values were grouped in 30 distances bins, such that each bin represented approximately the same number of pairwise LD estimates used to estimate  $N_e$  (Barbato et al., 2015). Within each bin the mean  $r^2$  was calculated and used to estimate  $N_t = 1/(4f(c)) (1/r^2 - 1)$ , where  $f(c) = c [(1 - c/2)/(1 - 2)^2]$  (Sved, 1971);  $c$  is the linkage distance inferred from the physical distance between SNPs assuming  $1 Mb \sim 1 cM$  and  $N_t$  represents the effective population size estimate at  $t = 1/2c$  generations ago.

**Signatures of Selection**

As the dataset analyzed here comprised *B. taurus* breeds from Europe and Africa, as well as *B. indicus* breeds from Asia and Africa, we sought to identify signatures of selection in a replicated manner (**Figure 1**). Initially we compared four taurine breeds (two European—Gascon and Chianina; and two African—Baoule and N'Dama) with four indicine breeds (two African – Zebu Fulani and Zebu Bororo; and the Asian Gir and Brazilian Nelore; Nelore was recently derived from Indian Ongole cattle, therefore, we broadly refer to these breeds as Asian indicine) to identify signatures of selection between *B. taurus* and *B. indicus*. Within *B. taurus*, we randomly chose three European breeds to compare against three African breeds in order to identify selection signatures specific to breeds in these continents. In the same way, for *B. indicus* we compared three Asian breeds against three African breeds (**Figure 1**). In contrast to previous studies, this experimental design enabled us identifying potential signatures of selection reflecting adaption to the local environment, instead of breed specific signatures that potentially reflect the breed's particular history.

The test of selection used was the cross population extended haplotype homozygosity test (XP-EHH) calculated for each SNP in the dataset as implemented in Selscan v1.0 (Szpiech and Hernandez, 2014). Selscan was used with default settings except for the maximum gap in bp allowed between SNPs, which was set to 400,000 to address the BovineSNP50 SNP density. For the comparisons between taurine and indicine breeds the following pairwise combinations were used: N'Dama (African taurine)/Gir (Asiatic indice), Gascone (European taurine)/Zebu Bororo (African indicine), Baoule (African taurine)/Zebu Fulani



**FIGURE 1 | Pairwise comparisons for analysis of selection.** Each box corresponds to one of the three types of pairwise comparisons carried out to detect signatures of selection. The blue box (comparison 1) indicates the comparison between four taurine and four indicine breeds used to detect signatures of selection between these breeds. The red box (comparison 2) shows the taurine breeds used in the comparison between African and not African breeds. The green box (comparison 3) shows the indicine breeds used in the comparison between African and not African breeds.

(African indicine) and Chianina (European taurine)/Nelore (Asian indicine). For the within-taurine comparison three pairwise tests were done between a European breed and an African breed, and for the within-indicine comparison three pairwise comparisons were carried out between an Asiatic breed and an African breed (**Figure 1**). For each pairwise comparison, the XP-EHH results were standardized for each chromosome separately, and the 5 and 95% quantiles of the standardized XP-EHH distribution for all markers were used as threshold to identify outlier SNPs, i.e., SNPs with XP-EHH values equal or smaller (more negative) or equal or larger (more positive) than the chosen thresholds. As multiple XP-EHH were carried out for each comparison (e.g., four comparisons between taurine and indicine breeds) this analysis was the equivalent of a replicated test. We took advantage of this design to identify SNPs under selection as those that showed for several pairwise comparisons

an extreme XP-EHH with the same sign (e.g., all pairwise comparisons should have an extreme positive XP-EHH).

For the between species comparison we defined the consistent SNPs under selection as those showing the same XP-EHH trend in at least three of the four pairwise comparisons. We did not restrict ourselves to only those SNPs where the four tests showed the same result to avoid excluding SNPs that failed to pass the 5% MAF threshold within a particular breed but which showed a consistent signature of selection across the remaining three comparisons. For the within species comparison of breeds occurring on different continents we identified consistent SNPs under selection as those showing an extreme XP-EHH with the same sign in each of the three comparisons made. The SNPs showing consistent signatures of selection across multiple pairwise comparisons were linked to neighboring genes using a window approach of 50 K base pairs, and a Gene Ontology analysis was carried out on these SNPs using Gorilla ([www.cblgorilla.cs.technion.ac.il](http://www.cblgorilla.cs.technion.ac.il); Eden et al., 2009).

## Results

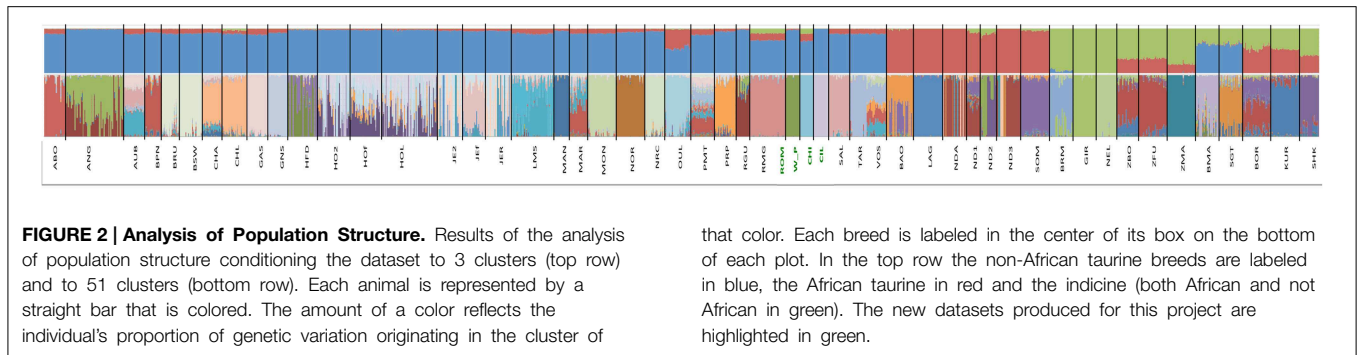
Of the 36,503 SNPs analyzed in 1345 animals representing taurine and indicine breeds, the observed heterozygosity varied between a minimum of 0.026 (Chillingham) and a maximum of 0.33 (the hybrid breed Beefmaster; **Table 1**). The comparison of expected heterozygosity (unbiased to sample size) between the taurine and indicine breeds was not significant, as well as the comparison between breeds in and out of Africa (all Welch *t*-test *p*-values > 0.2). However, the average observed heterozygosity in taurine breeds was slightly larger than that in indicine breeds [0.28 standard deviation (sd) 0.05 and 0.20 sd 0.04, respectively]. The average expected heterozygosity across breeds in both datasets was similar, i.e., 0.34 sd 0.0001. The inbreeding coefficient ( $F_{IS}$ ) was significantly lower for taurine breeds ( $F_{IS}$ : 0.18 sd 0.16) than for indicine breeds ( $F_{IS}$ : 0.42 sd 0.11; Welch *t*-test *p*-value: 0.0011). Within *B. taurus*, the African breeds presented a significantly higher inbreeding coefficient than the non-African taurines (Welch *t*-test *p*-value:  $1.54 \times 10^{-7}$ ). Similarly, the non-African taurines also presented significantly lower  $F_{IS}$  values than either of the groups of indicine breeds (Welch *t*-test *p*-value:  $1.84 \times 10^{-5}$  and 0.026, respectively). Among the non-African taurines, the Chillingham presented by far the highest inbreeding coefficient of the entire dataset ( $F_{IS}$ : 0.92), on average five-fold larger than other taurines, and two-fold larger than the indicine breeds. The Welsh White Park is thought to be related to the Chillingham breed, but presented an observed heterozygosity (0.245) only slightly smaller than the average value observed in the other non-African taurines. The  $F_{IS}$  value for Welsh White Park (0.27) was less than a third of that of Chillingham but still almost twice the average across the other non-African taurine breeds (0.14 sd 0.14). With the exception of Chillingham, the four breeds genotyped here presented typical observed heterozygosity and  $F_{IS}$  values for European taurine breeds (**Table 1**). Removing from the analysis the breeds with samples sizes smaller than 20 did not change the results (results not shown).

## Admixture and Genetic Relationships

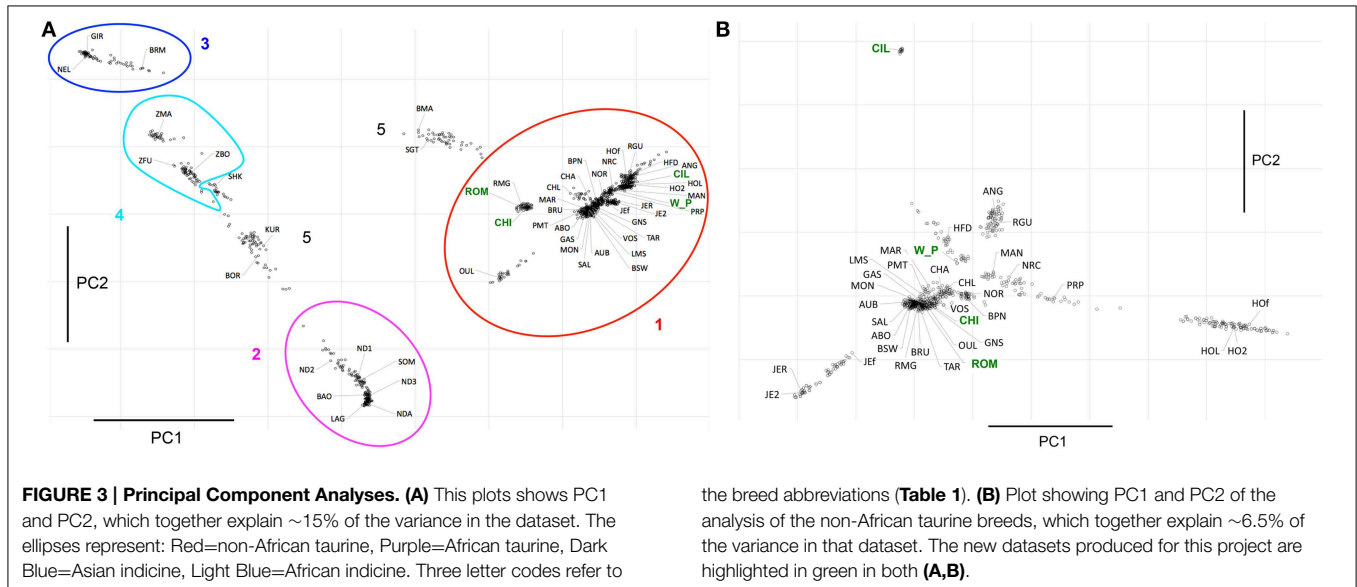
Admixture analysis between taurine and indicine breeds was initially run for values of the number of clusters (*K*) between 1 and 60. The cross validation (CV) statistic used to choose the most suitable number of clusters had its lowest value at *K* = 60. However, the shape of the CV curve suggested that higher *K* values may present lower CV. Consequently, we carried out runs of Admixture for larger values of *K* (e.g., 80). However, the CV values continued to reduce the larger the *K* numbers tested (SM Figure 1). While it may be possible that some family structure within breeds explains this trend, the substantially longer run time needed by the algorithm for such *K* values combined with the small decrease in CV suggests that clustering solutions larger than *K* ~60 potentially represent spurious results (a similar observation was made by Mctavish et al. (2013) who only achieved marginal increases on the likelihood scores using structure in their cattle dataset for values of *K* beyond 3).

The hierarchical clustering analysis with Admixture showed that the largest difference for *K* = 2 was between taurine and indicine breeds. Conditioning the data to *K* = 3 separated the taurine breeds with an African origin from the non-African, suggesting a large differentiation between taurine breeds in these two groups of samples (**Figure 2**). Contrastingly, for values up to *K* = 10, the African indicine breeds presented a remarkably similar distribution of genetic variation to that of the Asian indicine breeds, which did not form a separate cluster. Nonetheless, the African indicine breeds consistently presented a distinctive proportion of their genotype representing taurine ancestry throughout the analysis (a pattern almost entirely absent in the breeds of Asian origin). The *B. taurus*/*B. indicus* hybrids show a genetic make-up consisting of non-African taurine and *B. indicus* genetic components (Beefmaster and Santa Gertrudis), or African taurine and *B. indicus* (Borgou, Kuri, and Sheko).

The Admixture analysis was complemented with a principal component analysis (**Figure 3**). The first principal component (PC1) explained ~10% of the total variation in the data and separated taurine from indicine breeds. Besides of PC1, the only other component explaining a substantial part of the variance was PC2 (~5%), which separates the African taurine from the remaining taurine, as well as the African indicine from the Asian indicine breeds. PC3 to PC7 explained between less than 2 and 1.4% of the variance, while the remaining components explained less than 1% of the total variance, thus were not taken into consideration. PC3 separated Holstein from the remaining taurine breeds. PC4 identified a group of British breeds formed by Angus, Red Angus and Hereford, and PC5 identified a component of variation specific to Hereford separating it from the other taurines. PC6 identified the Chillingham cattle as well as the Jersey breed, and PC7 identified a specific component of variation to Chillingham separating it from the remaining taurine breeds. The combination of PC1 and PC2 identify the major groups of individuals described here in terms of species and geographic component (**Figure 3A**) with the hybrid breeds occurring between these. As expected from the Admixture analysis, the breeds Kuri, Borgou, and



that color. Each breed is labeled in the center of its box on the bottom of each plot. In the top row the non-African taurine breeds are labeled in blue, the African taurine in red and the indicine (both African and not African in green). The new datasets produced for this project are highlighted in green.



Sheko are placed with these principal components between the group of African taurine and indicine. Similarly, Beef Master and Santa Gertrudis are placed between the group of non-African taurine and indicine breeds, although from this analysis it is not clear whether their indicine genetic component is more similar to African indicine or Asian indicine.

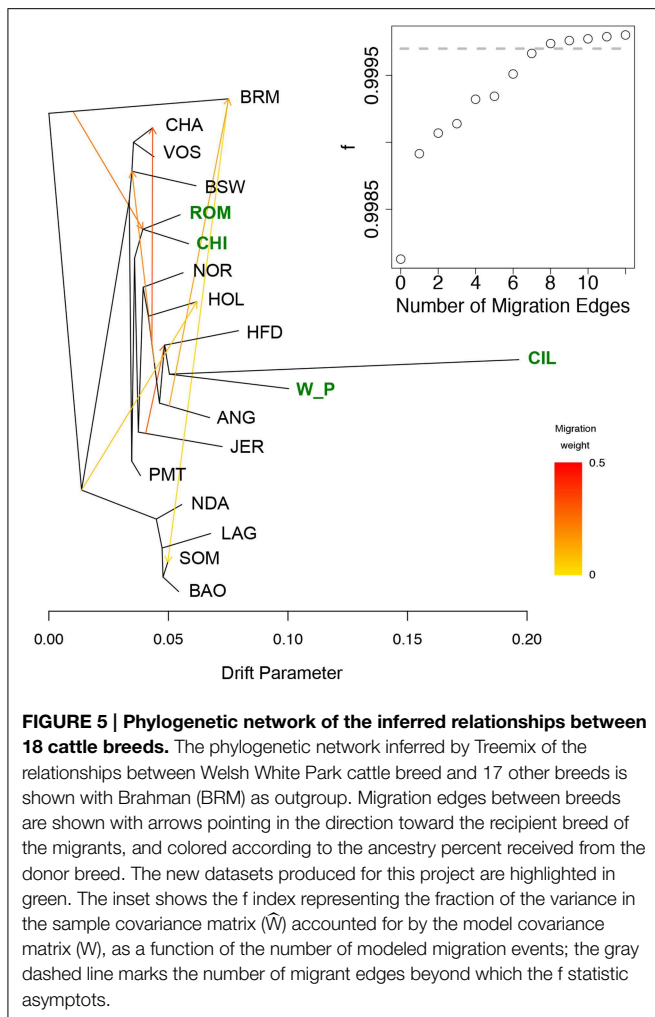
The methods described above identified differences between breeds but did not allow us to determine how these relate to each other. We therefore estimated a NeighbourNet network with these breeds using Reynold's distance. The torso of the network shows multiple pathways of connection between breeds reflecting the relatively recent divergence between many populations (Figure 4). However, the topology of the network resulted in the pattern expected from the previous analyses (Admixture and PCA), where a clear separation between taurine and indicine occurs, as well as the distinct separation between the African taurine breeds from the non-African taurines. Similarly to the PCA and in contrast to the Bayesian approach of Admixture, the network also resolved the African indicine breeds separately from their Asian counterparts. The dendrogram depicting the clustering of breeds using Reynold's distance recovered the grouping of breeds identified in the PCA with most branches

(83%) supported with high statistical support (bootstrap values higher than 70%; SM Figure 2). The clustering of the outgroups was not well resolved, except for the grouping between bison and yak. The indicine breeds clustered together with the Asian indicine showing a greater similarity to each other, and then to the Madagascan zebu. The two African indicine clustered together and then clustered with the remaining indicines. The taurine breeds separated into two clusters, one representing the African taurines and the other the non-African taurines. The hybrids Beefmaster and Santa Gertrudis clustered together within the non-African taurine group reflecting their larger taurine genetic component. Contrastingly, the African hybrids (Sheko, Borgou, and Kuri) clustered with the indicine breeds suggesting a higher indicine genetic component in their genotype.

A comparison between the Welsh White Park cattle and the other British breeds (Chillingham, Hereford, and Jersey) and Italian breeds (Chianina, Romagnola, Piemontese) was carried out to establish whether the UK white cattle have a genetic link with Roman white cattle, as is popularly believed. The Principal Components Analysis of this datasets identified the White Park breed clustering near the other British breeds (i.e., HFD, ANG, RGU), instead of closer to the Italian breeds (Figure 3B), as would be expected if White Park cattle had a strong Italian genetic





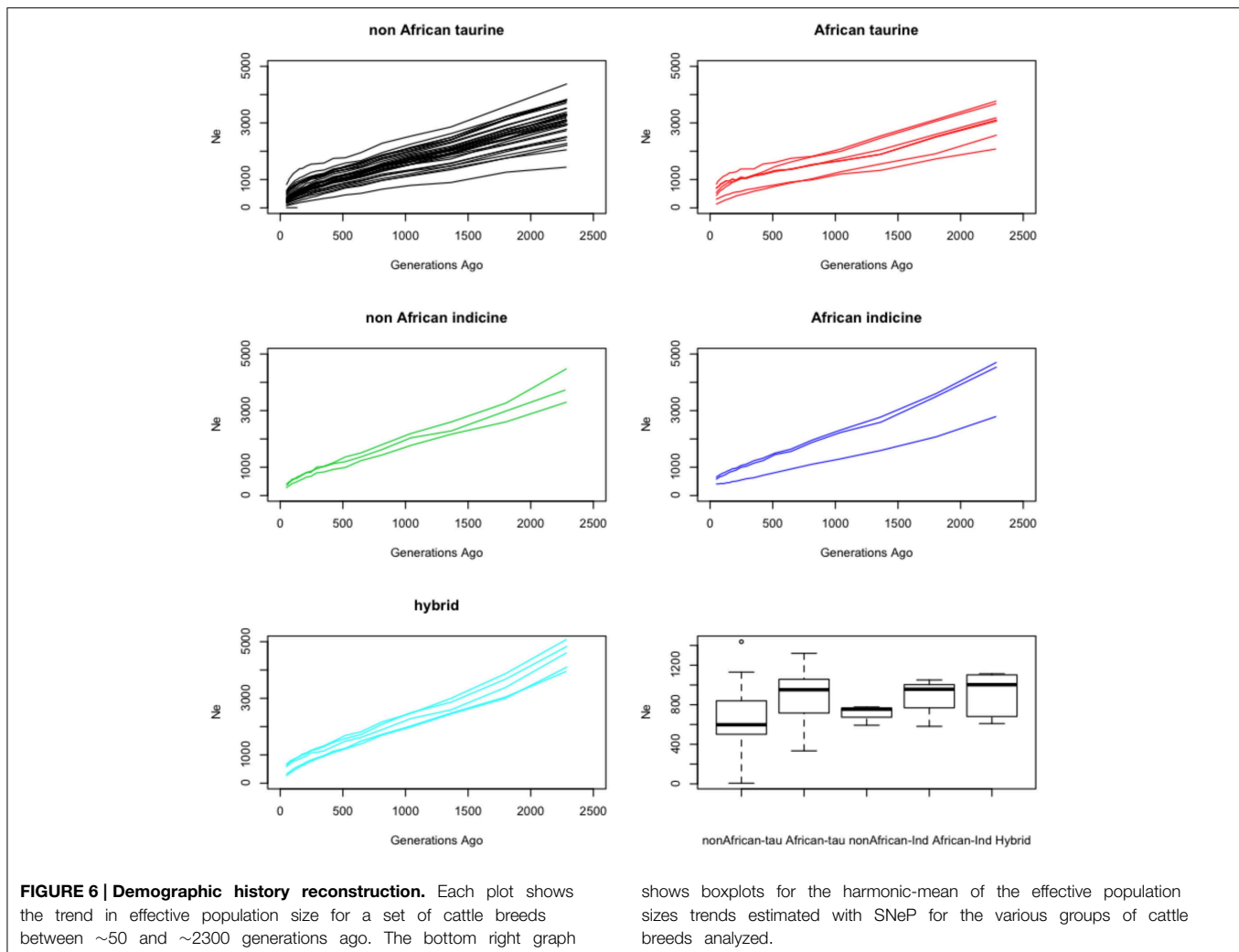


genes that occurred up to 50,000 bp downstream or upstream of the SNP position. No significantly enriched GO categories were found for any of the sets of breeds analyzed here. However, some patterns were observed, e.g., for the candidate SNPs under selection in Asian indicine several genes related to immunity were identified, e.g., LIPH (involved in platelet aggregation), *Tespa1* (involve in thymocyte development), *CCL14* (activates monocytes) (SM Table 1). However, interestingly, several RNA types were identified for several of the groups of breeds tested (Table 2). For the comparison between taurine and indicine breeds 6 snoRNA and 10 snRNA were identified among the taurine breeds, and two miRNA and 1 snoRNA in the indicine breeds. For the comparison between Asian and African indicine, a total of three miRNA, 1 snoRNA and 2 snRNA were identified for the Asian breeds and none for the African, while for the taurine comparison, the opposite was true, i.e., 1 snoRNA and 1 snRNA were identified for the African breeds and none for the European breeds. Overall this suggests that a substantial amount of selection signature occurred near non-coding RNAs that modify transcription of genes or of other RNAs.

## Discussion

This dataset represents one of the most comprehensive analyses of *B. taurus* and *B. indicus* cattle breed genetic data. Overall a higher observed genetic variation was observed in taurine breeds that originated or derived from European cattle breeds ( $\sim 0.29$ ), than in the other cattle breeds. Both African taurine and indicine breeds presented an intermediate level of observed heterozygosity ( $\sim 0.22$  in both sets) and the Asian indicine breeds presented the lowest observed heterozygosity ( $\sim 0.17$ ). While the observed difference could have a generic demographic explanation, e.g., a larger effective population size in the taurine breeds of European descent, such pattern is not consistent with what has been described with other types of molecular markers such as microsatellites (Rothammer et al., 2013). The most likely explanation for this pattern is ascertainment bias (Gautier et al., 2009; Matukumalli et al., 2009). In particular, for those breeds not included in the panel of breeds used to identify SNPs in the BovineSNP50 chip assay, minimum allele frequencies are expected to be substantially lower and the genetic variation captured with the chosen SNPs may not be reflective of the true underlying genetic variation in those breeds. In this case, SNP discovery was based on the genome sequence reads of five taurine breeds (Holstein, Angus, Limousin, Jersey, Norwegian Red) and one indicine (Brahman) breed compared to the reference genome of a Hereford cow (Bovine Hapmap et al., 2009). Ascertainment bias deriving from the design of this SNP array was previously described for a subset of the dataset analyzed here, where up to 30% of the SNPs had a  $MAF > 0.3$  in taurine breeds, while only 19% of the SNPs had an equivalent  $MAF$  in indicine breeds (Nelore, Brahman, and Gir) (Bovine Hapmap et al., 2009). Similarly, the analysis of Holstein, Angus and Brahman samples carried out with a subset of the SNPs in the BovineSNP50 chip assay ascertained in each of the breeds separately, showed how Brahman animals presented a lower heterozygosity using the taurine SNPs ( $p$ -value  $< 0.05$  and  $p$ -value = 0.07, respectively) than using the indicine SNPs (Neto and Barendse, 2010). Contrastingly, in the same analysis, for the Holstein and Angus derived SNPs, the taurine samples presented an observed heterozygosity typical to that of any other taurine breed. Furthermore, as no African taurine breeds and most indicine breeds were not included in the SNP ascertainment, the observed levels of genetic variation are likely to be an underestimate of the true genetic variation in these breeds (Matukumalli et al., 2009), and the associated inbreeding coefficient may consequently be an overestimate of the real ones. Thus, comparisons of genetic diversity discussed here are only likely to be meaningful within major genetic groups.

Among all cattle, the British Chillingham breed possessed unusually low observed heterozygosity (0.026) and an extremely high inbreeding coefficient (0.92). The Chillingham occurs in Northumberland, UK, and for the last  $\sim 300$  years has lived in a feral state with almost no human intervention and no immigration (Visscher et al., 2001). Additionally, this breed passed through a bottleneck in 1947 where the herd reduced to five males and eight females. Previous analysis of genetic variation using 25 microsatellites identified only one marker with



polymorphism (Visscher et al., 2001). Additionally, sequences of the full mitochondria of eight animals (sequencing depth ~2935x using next generation sequencing) revealed a single haplotype in the herd that differentiated them from other taurine breeds by three mutations, and one mutation (likely to be a recurrent mutation) that related them to indicine breeds and yak (Hudson et al., 2012). These studies suggest that Chillingham probably survives because the population has purged most deleterious mutations. The results shown here represent the largest study of genetic variation in this breed so far and did find similarly low levels of genetic variation in the breed (~900 polymorphic SNPs out of ~35,000). An extended analysis including more individuals and genetic markers in this breed may identify genetic variation of evolutionary importance not captured in this study, as well as shedding light on the distribution of homozygous and heterozygous tracts in the breed that could be used to monitor its genetic health.

### Population Structure

The analyses of population structure easily identified the taurine and indicine genetic components in the dataset (Figure 2).

However, when conditioning the number of clusters in the data to  $K = 2$ , the African taurine cattle presented a genetic composition similar to that of the American hybrid breeds Beef Master and Santa Gertrudis, the latter two are breeds of recent origin (less than 100 years old). Increasing the number of clusters to  $K = 3$  clearly separates the African taurines in their own cluster, suggesting that despite the strong non-African taurine and indicine genetic component (as seen for  $K = 2$ ), African taurines (both longhorn and shorthorn) have a substantial third genetic component specific to Africa which may derive from cross breeding with local aurochs (Mctavish et al., 2013). Consistent with previous observations, the African taurine N'Dama (ND1 and ND2) and Somba (SOM) breeds presented a minor indicine genetic component (Gautier et al., 2010). Among the hybrids, Beefmaster (BMA) and Santa Gertrudis (STG) presented a major genetic component of non-African taurine origin and the remaining of indicine origin, in accordance with their breed formation. Beefmaster and Santa Gertrudis are both hybrid breeds of North American origin resulting from the cross between European taurine cattle of British origin (Hereford and Shorthorn) and Brahman cattle (Mctavish et al.,

**TABLE 2 | Classes of molecules under selection.**

Molecule Class	Taurine vs. Indicine	
	Taurine	Indicine
Protein Coding	15	14
RNA (mi, sno, sn)	0,6,10	2,1,0
rRNA	0	0
Pseudogene	1	1
	Indicine	
	Asian	African
Protein coding	46	2
RNA (mi, sno, sn)	3,1,2	0
rRNA	3	0
Pseudogene	3	0
	Taurine	
	European	African
Protein coding	5	1
RNA (mi, sno, sn)	0	0,1,1
rRNA	0	0
pseudogene	0	0

DNA sequences coding for molecules in close proximity to SNPs under selection were grouped as Protein Coding, RNA types (mi, sno, and sn), rRNA and pseudogenes. The number of SNPs under selection close to each of these classes is shown for each of the three analyses presented.

2013). On the other hand, the remaining three hybrid breeds, the Borgou, Kuri, and Sheko present a genetic make-up of African taurine and indicine origin, with the first two having a somewhat larger taurine proportion than the Sheko breed. Among the taurine breeds, the Moroccan Oulmès Zaer (OUL) breed of unknown origin, presented approximately  $\sim 3/5$  of its genotype of non-African taurine type and the remaining of African taurine type (**Figure 2**), consistent with previous results suggesting an influence of African longhorn taurine into the otherwise European taurine genetic background of this breed (Gautier et al., 2010; Decker et al., 2014). Interestingly, the non-African taurine cattle presented around 10% or less of their genetic make-up showing African taurine ancestry (**Figure 2**), possibly deriving from cross Mediterranean movement of cattle during the last millennia (Dürst, 1899; Do Vale, 1907; Felius et al., 2011). A further indicine genetic component was also detected in the French Charolais and Gascon breeds (<5%), and the Italian Romagnola and Chianina breeds ( $\sim 10\%$ ). These results support observations by Decker et al. (2014) that suggest indicine admixture in European taurine breeds (the bulk of non-African taurines in this study) is rare instead of widespread (Mctavish et al., 2013). The genetic variation of African taurine is clearly distinct from that of other cattle breeds, and the PCA analysis suggests that the three main cattle groups (African and non-African taurine, and indicine) are almost as different from each other. While these results do not provide evidence for or against the hypothesis of a third domestication event in Africa (Felius et al., 2011), it suggests that African taurine cattle present a unique characteristics in their gene-pool.

The analysis of population structure using Admixture resulted in asymptoting CV values reflecting problems in identifying

the true number of clusters in the data, possibly due to the presence of family structure within the breeds, or relatively low divergence between the estimated clusters. Thus, we also carried out principal components analysis, and as expected from the problems with Admixture, there is relatively low variation among cattle groups, i.e., the largest principal component (PC1) only explains  $\sim 10\%$  of the variance in the dataset (separating taurine from indicine breeds), while PC2 explains  $\sim 5\%$  of the variance (separating African from non-African taurines). Consistent with the Admixture result, the combination of PC1 and PC2 places the hybrid breeds Kuri, Borgou and Sheko between the African taurine and African indicine clusters, with the Sheko much closer to the indicine cluster reflecting its larger indicine background than Kuri and Borgou. The American Beef Master and Santa Gertrudis are placed between the non-African taurine and Asian indicine, but somewhat closer to the taurine cluster reflecting the larger taurine component of their genetic variation.

Lastly, the NeighbourNet analysis and the dendrogram identify the Madagascan zebu as more similar to the Asian indicine than to the other two African indicine. While this placing may be biased due to the ascertainment bias in the SNP array, it is tempting to suggest that the Madagascan population share a more recent history with Asian indicine breeds than the continental African breeds (e.g., a more recent establishment of the Madagascan population).

### Roman Origin of Welsh White Park

Within the United Kingdom, traditional knowledge suggests that the Welsh White Park were brought to Britain by the Romans (Felius et al., 2011). While there is no evidence of the Romans having introduced cattle to Britain (Visscher et al., 2001), PCA places the White Park almost in the middle of the group of non-African taurine breeds (**Figure 3A**), consistently with the results by Decker et al. (2014) based on genotypes from five individuals. In particular, a comparison between Welsh White Park and the three Italian breeds in this dataset, Chianina, Romagnola and Piedmontese, revealed that the Italian breeds are more closely related to each other than to White Park, and the Welsh breed is more closely related to the other British breeds (**Figures 3B, 5**). Nevertheless, while this suggests that Welsh White Park are not similar to the Italian breeds it still does not answer the question regarding their origin. Both the NeighbourNet (**Figure 4**) and the phylogenetic network (**Figure 5**) show that the Welsh White Park and Chillingham cluster near each other (**Figure 4**). Such observation could be the outcome of genetic drift or inbreeding (or an interplay of these two parameters) in these breeds driving the divergence of their allelic frequencies (particularly in the case of Chillingham). Consistently with this hypothesis, the phylogenetic network analysis modeling genetic drift suggests that Welsh White Park and Chillingham have been through substantial drift in comparison to other taurines (**Figure 5**). Additionally, the Treemix analysis suggests that no migration events connect either the Welsh White Park or Chillingham with any of the Italian breeds, unless an additional edge that marginally increases the  $f$ -statistic by 0.01% (i.e., from 99.96 to 99.97% of the variance explained) is added; however, that additional migration edge is from Welsh White

Park into Romagnola (contrary to common believe of an Italian ancestry of White Park) and has a weight near 0 (SM Figure 4) indicating that a very small fraction (if at all) of the Romagnola allelic variation may be of Welsh origin. While such idea could be appealing to some, it does not necessarily reflect a real historic event as the inferred edge provides an almost negligible contribution to the variance in the phylogram and corresponds to a minimal contribution of Welsh allelic variants into Romagnola. Additionally, the question of whether the inferred patterns of migration by Treemix are somehow dependent on the data analyzed could be raised. Further studies on the genetic history of these iconic breeds should include additional breeds such as the Podolian Hungarian Gray, or breeds close to the domestication center in order to better characterize the origin of this animals.

## Demographic History

The demographic history of a population can be traced using the distribution of LD across the genome (e.g., Hill and Robertson, 1968), and if the cattle domestication event was marked by a strong bottleneck, it is expected that such a pattern should be visible with this dataset. However, it is important to note that the analyses presented here might provide a biased estimate of the true extent of demographic history, as LD calculated over SNP array markers is expected to decay at a lower rate than when the same estimates are obtained from whole-genome sequence data (Qanbari et al., 2014).

All the breeds analyzed present trends marked by a decline in their effective population size during the last ~2000 generations (Figure 6). Assuming an average generation length of 4 years, that suggests a constant decline across all breeds since the onset of the domestication process approximately 8000 years ago (Gautier et al., 2010). These results are consistent with observations describing a general decrease in breeds such as Angus, Holstein, Hereford, N'Dama and Brahman over the same period of time (Gautier et al., 2007; De Roos et al., 2008; Flury et al., 2010). Interestingly, overall, the harmonic-mean effective population size of the non-African taurine breeds was smaller than that of the non-African indicine, but both of those means were much smaller than that of the African taurines and indicines (Figure 6). These results suggest that either crossing between African taurine and indicine breeds has increased the effective population size in these breeds by moving genetic variation between breeds, and/or that more benign management practices in Africa have resulted in a milder reduction in effective population size. Although the second hypothesis cannot be ruled out, consistent with the first hypothesis, the hybrid breeds present overall a higher average effective population size.

## Signatures of Selection

The unique nature of the data presented here provided us with a potentially powerful statistical framework to detect signatures of selection. In contrast to other studies, we examined multiple breeds for each species, accounting for their distribution in and out of Africa to search for signatures of selection in the equivalent of a replicated experimental design, where instead of identifying selection in a particular breed, we used several breeds of each species in each continent as biological replicas. The test used is

an extension of Sabeti's extended haplotype homozygosity (Sabeti et al., 2002), which is based on the principle that a site under positive selection will rapidly increase in frequency, bringing neighboring variants to a high frequency in a process called hitchhiking (Smith and Haigh, 1974), thereby removing genetic variation around the site under selection (a selective sweep) and leaving a track of extended homozygosity. The magnitude of the effect of the selected site on the neighboring variation largely depends on how strong and recent the selective pressure is. If the process is slow, because selection is weak, recombination events separate the selected site from neighboring variants breaking down linkage and consequently resulting in a short area of extended homozygosity (Sabeti et al., 2002). On the contrary, if the process is fast because selection is strong, the area of extended homozygosity will be large, as there is not enough time for recombination to break down linkage. Consequently, identifying old signatures of positive selection using the EHH or any of its derived versions (e.g., XP-EHH) is rather difficult as the expected pattern of linkage disequilibrium is likely to have disappeared (Sabeti et al., 2007; Chen et al., 2010). However, despite the ancient divergence between *B. taurus* and *B. indicus*, i.e., approximately 250,000 years old (Bradley et al., 1996; Bovine Hapmap et al., 2009), a consistent signature of selection was found for several thousands of SNPs (5385 SNPs) probably reflecting relatively recent selective events rather than ancient ones.

Consistent with the expectation for the comparison between breeds in both species, a similar number of SNPs in each species showed signatures of selection (i.e., 3029 in taurine and 2385 in indicine). For the African vs. non-African comparison of taurine breeds 1872 of the SNPs under selection were on European breeds and 1739 were in African breeds. In contrast, for the comparison of African vs. non-African indicine breeds, 1768 of the SNPs under selection were on the Asian indicine breeds and 2049 were in the African breeds. The difference between the comparison of taurine breeds and indicine breeds in and out of Africa may reflect the shorter period of time that the later have spent in Africa since their introduction to that continent. The first domestic cattle to have entered Africa were taurines approximately 7000 years ago (Payne and Hodges, 1997; Hassan, 2000; Felius et al., 2014). These domestic animals probably entered via a route crossing through Egypt and moving into northern Africa. In contrast, indicine breeds entered Africa much later via routes leading to East Africa via Egypt some time between ~3500 and 2500 years ago (Payne and Hodges, 1997; Hanotte et al., 2002). The timing of these migration events into Africa implies that taurine breeds have been in Africa between two and three times longer than indicine breeds.

At least three non-exclusive hypotheses can be derived from the taurine colonization history of Africa. Firstly, African taurine have become clearly differentiated from other non-African taurines as genetic drift exacerbated differences between the taurine propagules that spread from the domestication center into Europe and Africa. Secondly, African taurines had a longer time to hybridize with local aurochs, as possibly was the case for the European taurines, which hybridized with European aurochs (Decker et al., 2014). Consequently, African taurines are expected

to carry unique genetic variation (deriving from local aurochs) that differentiates them from non-African taurines, as African aurochs were likely to genetically differ from European aurochs due to isolation by distance (Perez-Pardal et al., 2010; Mctavish et al., 2013). Thirdly, by having spent a longer time in Africa, African taurines are expected to have had a longer time to adapt to the local environment and although some of their signatures of local adaptation may have dissipated, they may have left a trace of genetic identity that helps differentiating African taurine from other cattle breeds. In contrast, *B. indicus* was introduced to Africa approximately ~2000 years ago, and during this short period of time African indicine breeds may have not had enough time to differentiate from other indicine breeds, as much as the African taurines did. However, as the presence of indicine breeds in Africa is recent it is reasonable to expect that these carry more signatures of selection than African taurines (~22 vs. ~17%, respectively).

The analysis of selection carried out here sought to identify SNPs under selection using the equivalent of a replicated experiment. For that purpose it was necessary that several pairwise comparisons identified the same SNP under selection showing the same sign in the XP-EHH analysis (e.g., a SNP showing selection for the European taurine variant in the three pairwise comparisons). The rationale for this is that genetic drift can cause extreme differences in allele frequencies between populations, however, the chance that the same allele changes allele frequency in the same direction in multiple populations is low (e.g., A gets fixed by chance in the three replicates). While our approach is expected to be robust as it combines the power of the track of linkage disequilibrium left after positive selection and its replicability across populations occupying similar environments (e.g., three European breeds tested against three African breeds), it is possible that weaker signatures of selection were not identified. Nevertheless, it is interesting to note that for the pairwise comparisons where several genes were identified to be in close proximity to the SNPs under selection some chromosomes seems to be more frequent under selection than others. For example, among the Asian indicine, eight genes occurred on the first chromosome, and 11 on the 19th chromosome (SM Table 1), while three out of the four genes identified among the European taurine breeds occurred on the 10th chromosome. Lastly, it is expected that short term adaptive processes may be mediated through modifications of regulatory pathways (Karim et al., 2011; Visscher and Goddard, 2011; David et al., 2013; Furusawa and Kaneko, 2013). Consistently with such hypothesis we find that several SNPs under selection occur in close proximity to RNA coding sequences involved in transcription regulation (e.g., miRNAs) and modification of other RNAs (e.g., snoRNAs).

## References

- Abraham, G., and Inouye, M. (2014). Fast principal component analysis of large-scale genome-wide data. *PLoS ONE* 9:e93766. doi: 10.1371/journal.pone.0093766
- Ajmone-Marsan, P., Garcia, J. F., and Lenstra, J. (2010). On the origin of cattle: how aurochs became cattle and colonized the world. *Evol. Anthropol.* 19, 148–157. doi: 10.1002/evan.20267
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Barbato, M., Orozco-Terwengel, P., Tapio, M., and Bruford, M. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* 6:109. doi: 10.3389/fgene.2015.00109
- Beja-Pereira, A., Caramelli, D., Lalueza-Fox, C., Vernesi, C., Ferrand, N., Casoli, A., et al. (2006). The origin of European cattle: evidence from
- The domestication of cattle in Eurasia had profound implications for society (Ajmone-Marsan et al., 2010; Groeneveld et al., 2010; Felius et al., 2014). Cattle are one of the major livestock species in the world and represent a large assemblage of widespread breeds as well as locally adapted breeds (Ajmone-Marsan et al., 2010; Felius et al., 2011, 2014). Studying the genetic resource of cattle breeds around the world is an on going work (e.g., Bovine Hapmap et al., 2009; Gautier et al., 2010; Mctavish et al., 2013; Decker et al., 2014) and approaches like the ones used here show the power that multiple comparisons have to derive particular trends and patterns that have characterized these breeds' histories. Additionally, identifying gaps representing regions where cattle has not yet been studied, as well as breeds that have not yet been genetically characterized is of utter importance to identify the genetic variation with adaptive value (e.g., local adaptation) and that should be the target of conservation efforts. Finally, whether Welsh White Park cattle were introduced to Britain by the Romans seems unlikely with the data described here, but cannot be entirely ruled out. The current results suggest that the breed presents a distinct genetic variation possible reflecting its ancient origin, but without some sort of calibration for the molecular phylogeny, an Approximate Bayesian computation approach that estimates times of divergence between breeds, or its comparison to local breeds close to the domestication center, this question is left partially unanswered (Csillery et al., 2010a,b; Sunnaker et al., 2013).

## Author Contributions

PO and MWB conceived the project; PO, MB, EN, and FB carried out analyses; WD, DW, MM, AS, PA contributed samples and reagents; PO wrote the manuscript; MB, EN, FB, and MWB revised the manuscript. All authors approved the final manuscript.

## Acknowledgments

The authors acknowledge the Chillingham Wild Cattle Association for the collection of samples for DNA analysis, as well as the NextGen consortium (<http://nextgen.epfl.ch/>) and Genomic-Resources (European Science Foundation; <http://www.esf.org/index.php?id=7009>) for funding, and two reviewers for their comments.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fgene.2015.00191/abstract>

- modern and ancient DNA. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8113–8118. doi: 10.1073/pnas.0509210103
- Biffani, S., Marusi, M., Biscarini, F., and Canavesi, F. (2005). Developing a genetic evaluation for fertility using angularity and milk yield as correlated traits. *Interbull. Bull.* 33, 63–66.
- Bollongino, R., Elsner, J., Vigne, J. D., and Burger, J. (2008). Y-SNPs do not indicate hybridisation between European aurochs and domestic cattle. *PLoS ONE* 3:e3418. doi: 10.1371/journal.pone.0003418
- Bovine Hapmap, C., Gibbs, R. A., Taylor, J. F., Van Tassell, C. P., Barendse, W., Eversole, K. A., et al. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936
- Bradley, D. G., Machugh, D. E., Cunningham, P., and Loftus, R. T. (1996). Mitochondrial diversity and the origins of African and European cattle. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5131–5135. doi: 10.1073/pnas.93.10.5131
- Bulmer, M. G. (1971). The effect of selection on genetic variability. *Am. Nat.* 105, 201–211. doi: 10.1086/282718
- Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. doi: 10.1101/gr.100545.109
- Cobb, M. (2006). Heredity before genetics: a history. *Nat. Rev. Genet.* 7, 953–958. doi: 10.1038/nrg1948
- Corbin, L. J., Liu, A. Y., Bishop, S. C., and Woolliams, J. A. (2012). Estimation of historical effective population size using linkage disequilibria with marker data. *J. Anim. Breed. Genet.* 129, 257–270. doi: 10.1111/j.1439-0388.2012.01003.x
- Csillery, K., Blum, M. G. B., Gaggiotti, O. E., and Francois, O. (2010b). Invalid arguments against ABC: reply to AR Templeton. *Trends Ecol. Evol.* 25, 490–491. doi: 10.1016/j.tree.2010.06.011
- Csillery, K., Blum, M. G., Gaggiotti, O. E., and Francois, O. (2010a). Approximate Bayesian Computation (ABC) in practice. *Trends Ecol. Evol.* 25, 410–418. doi: 10.1016/j.tree.2010.04.001
- David, L., Ben-Harosh, Y., Stolovicki, E., Moore, L. S., Nguyen, M., Tamse, R., et al. (2013). Multiple genomic changes associated with reorganization of gene regulation and adaptation in yeast. *Mol. Biol. Evol.* 30, 1514–1526. doi: 10.1093/molbev/mst071
- Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcala, A., Sonstegard, T. S., et al. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* 10:e1004254. doi: 10.1371/journal.pgen.1004254
- De Roos, A. P., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179, 1503–1512. doi: 10.1534/genetics.107.084301
- Do Vale, M. (1907). *Bovideos Portuguezes* Lisbon: La Bécarre.
- Dürst, J. (1899). *Die Rinder von Babylonien, Assyrien un Ägypten und ihr Zusammenhang mit den Rindern der Alten Welt*. Berlin: Gerog Reimber. doi: 10.5962/bhl.title.86398
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48. doi: 10.1186/1471-2105-10-48
- Edwards, C. J., Ginja, C., Kantanen, J., Perez-Pardal, L., Tresset, A., Stock, F., et al. (2011). Dual origins of dairy cattle farming - evidence from a comprehensive survey of european Y-Chromosomal Variation. *PLoS ONE* 6:e15922. doi: 10.1371/journal.pone.0015922
- Fao. (2007). "Global plan of action for animal genetic resources and the Interlaken Declaration," in *International Technical Conference on Animal Genetic Resources for Food and Agriculture* (Rome).
- Felius, M., Beerling, M., Buchanan, D., Theunissen, B., Koolmees, P., and Lenstra, J. (2014). On the history of cattle genetic resources. *Diversity* 6, 705–750. doi: 10.3390/d6040705
- Felius, M., Koolmees, P., Theunissen, B., Consortium, E. C. G. D., and Lenstra, J. (2011). On the breeds of cattle - historic and current classifications. *Diversity* 3, 660–692. doi: 10.3390/d3040660
- Felsenstein, J. (1993). *PHYMLIP (Phylogeny Inference Package) version 3.5c*. 3.69 ed. (Seattle, WA: Department of Genetics, University of Washington).
- Flori, L., Gonzatti, M. I., Thevenon, S., Chantal, I., Pinto, J., Berthier, D., et al. (2012). A quasi-exclusive European ancestry in the Senepol tropical cattle breed highlights the importance of the slick locus in tropical adaptation. *PLoS ONE* 7:e36133. doi: 10.1371/journal.pone.0036133
- Flury, C., Tapio, M., Sonstegard, T. S., Drögemüller, T., Leeb, H., Simianer, H., et al. (2010). Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *J. Anim. Breed. Genet.* 127, 339–347. doi: 10.1111/j.1439-0388.2010.00862.x
- Furusawa, C., and Kaneko, K. (2013). Epigenetic feedback regulation accelerates adaptation and evolution. *PLoS ONE* 8:e61251. doi: 10.1371/journal.pone.0061251
- 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. doi: 10.1534/genetics.107.075804
- 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. doi: 10.1186/1471-2164-10-550
- Gautier, M., Laloe, D., and Moazami-Goudarzi, K. (2010). Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS One* 5:e13038. doi: 10.1371/journal.pone.0013038
- Gotherstrom, A., Anderung, C., Hellborg, L., Elburg, R., Smith, C., Bradley, D. G., et al. (2005). Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proc. Biol. Sci.* 272, 2345–2350. doi: 10.1098/rspb.2005.3243
- Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., et al. (2010). Genetic diversity in farm animals—a review. *Anim. Genet.* 41(Suppl. 1), 6–31. doi: 10.1111/j.1365-2052.2010.02038.x
- Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee, Y., Hill, E. W., and Rege, J. E. (2002). African pastoralism: genetic imprints of origins and migrations. *Science* 296, 336–339. doi: 10.1126/science.1069878
- Hassan, F. (2000). *The Origins and Development of African Livestock: Archaeology, Genetics, Linguistics and Ethnography*. London: University College London Press.
- Herrero, M., and Thornton, P. K. (2013). Livestock and global change: emerging issues for sustainable food systems. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20878–20881. doi: 10.1073/pnas.1321844111
- Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226–231. doi: 10.1007/BF01245622
- Hudson, G., Wilson, I., Payne, B. I., Elson, J., Samuels, D. C., Santibanez-Korev, M., et al. (2012). Unique mitochondrial DNA in highly inbred feral cattle. *Mitochondrion* 12, 438–440. doi: 10.1016/j.mito.2012.05.003
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. doi: 10.1093/molbev/msj030
- Kantanen, J., Olsaker, I., Holm, L. E., Lien, S., Vilkkilä, J., Brusgaard, K., et al. (2000). Genetic diversity and population structure of 20 North European cattle breeds. *J. Hered.* 91, 446–457. doi: 10.1093/jhered/91.6.446
- Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J. A., Baurain, D., et al. (2011). Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nat. Genet.* 43, 405–413. doi: 10.1038/ng.814
- Lewis, J., Abas, Z., Dadousis, C., Lykidi, D., Paschou, P., and Drineas, P. (2011). Tracing cattle breeds with principal components analysis ancestry informative SNPs. *PLoS ONE* 6:e18007. doi: 10.1371/journal.pone.0018007
- Loftus, R. T., Machugh, D. E., Bradley, D. G., Sharp, P. M., and Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2757–2761. doi: 10.1073/pnas.91.7.2757
- Machugh, D. E., Loftus, R. T., Cunningham, P., and Bradley, D. G. (1998). Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Anim. Genet.* 29, 333–340. doi: 10.1046/j.1365-2052.1998.295330.x
- Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., et al. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4:e5350. doi: 10.1371/journal.pone.0005350
- Mctavish, E. J., Decker, J. E., Schnabel, R. D., Taylor, J. F., and Hillis, D. M. (2013). New World cattle show ancestry from multiple independent domestication events. *Proc. Natl. Acad. Sci. U.S.A.* 110, E1398–1406. doi: 10.1073/pnas.1303367110
- Medjugorac, I., Kustermann, W., Lazar, P., Russ, I., and Pirchner, F. (1994). Marker-derived phylogeny of european cattle supports domestic expansion of agriculture. *Anim. Genet.* 25, 19–27. doi: 10.1111/j.1365-2052.1994.tb00399.x
- Neto, L. R. P., and Barendse, W. (2010). Effect of SNP origin on analyses of genetic diversity in cattle. *Anim. Product. Sci.* 50, 792–800. doi: 10.1071/AN10073

- Payne, W., and Hodges, J. (1997). *Tropical Cattle, Origins, Breeds and Breeding Policies*. Oxford: Blackwell Science Ltd.
- Pellecchia, M., Negrini, R., Colli, L., Patrini, M., Milanese, E., Achilli, A., et al. (2007). The mystery of Etruscan origins: novel clues from *Bos taurus* mitochondrial DNA. *Proc. Biol. Sci.* 274, 1175–1179. doi: 10.1098/rspb.2006.0258
- Perez-Pardal, L., Royo, L. J., Beja-Pereira, A., Curik, I., Traore, A., Fernandez, I., et al. (2010). Y-specific microsatellites reveal an African subfamily in taurine (*Bos taurus*) cattle. *Anim. Genet.* 41, 232–241. doi: 10.1111/j.1365-2052.2009.01988.x
- Perry, B. D., Grace, D., and Sones, K. (2013). Current drivers and future directions of global livestock disease dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20871–20877. doi: 10.1073/pnas.10129531108
- Pickrell, J. K., and Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967. doi: 10.1371/journal.pgen.1002967
- Pryce, J., Royal, M., Garnsworthy, P., and Maod, I. (2004). Fertility in high-producing dairy cow. *Livestock Product. Sci.* 86, 125–135. doi: 10.1016/S0301-6226(03)00145-3
- Pryce, J., Veerkamp, R., Thompson, R., Hill, W. G., and Simm, G. (1997). Genetic aspects of common health disorders and measures of fertility in Holstein Friesian dairy cattle. *Anim. Sci.* 65, 353–360. doi: 10.1017/S135772980008559
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., et al. (2014). Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 10:e1004148. doi: 10.1371/journal.pgen.1004148
- Rischkowsky, B., and Pilling, D. (2007). *The State of the World's Animal Genetic Resources for Food and Agriculture*. Rome: Food and Agriculture Organization of the United Nations.
- Rothhammer, S., Seichter, D., Forster, M., and Medugorac, I. (2013). A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics* 14:908. doi: 10.1186/1471-2164-14-908
- Russell, N. (2007). *Like Engend'ring Like. Heredity and Animal Breeding in Early Modern England*. Cambridge, MA: Cambridge University Press.
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837. doi: 10.1038/nature01140
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* 449, 913–918. doi: 10.1038/nature06250
- Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644. doi: 10.1086/502802
- Simm, G. (1998). *Genetic Improvement of Cattle and Sheep*. Wallingford: Farming Press.
- Smith, J. M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet. Res.* 23, 23–35. doi: 10.1017/S0016672300014634
- Sunnaker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., and Dessimoz, C. (2013). Approximate Bayesian computation. *PLoS Comput. Biol.* 9:e1002803. doi: 10.1371/journal.pcbi.1002803
- Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Popul. Biol.* 2, 125–141. doi: 10.1016/0040-5809(71)90011-6
- Sved, J. A., Mcrae, A. F., and Visscher, P. M. (2008). Divergence between human populations estimated from linkage disequilibrium. *Am. J. Hum. Genet.* 83, 737–743. doi: 10.1016/j.ajhg.2008.10.019
- Szpiech, Z. A., and Hernandez, R. D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* 31, 2824–2827. doi: 10.1093/molbev/msu211
- Taberlet, P., Coissac, E., Pansu, J., and Pompanon, F. (2011). Conservation genetics of cattle, sheep, and goats. *C. R. Biol.* 334, 247–254. doi: 10.1016/j.crv.2010.12.007
- Thornton, P. K. (2010). Livestock production: recent trends, future prospects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2853–2867. doi: 10.1098/rstb.2010.0134
- Troy, C. S., Machugh, D. E., Bailey, J. F., Magee, D. A., Loftus, R. T., Cunningham, P., et al. (2001). Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410, 1088–1091. doi: 10.1038/35074088
- Visscher, P. M., and Goddard, M. E. (2011). Cattle gain stature. *Nat. Genet.* 43, 397–398. doi: 10.1038/ng.819
- Visscher, P. M., Smith, D., Hall, S. J., and Williams, J. L. (2001). A viable herd of genetically uniform cattle. *Nature* 409, 303. doi: 10.1038/35053160
- Wendorf, F., and Schild, R. (1994). Are the early holocene cattle in the eastern Sahara domestic or wild? *Evol. Anthropol.* 3, 118–128. doi: 10.1002/evan.1360030406
- Wood, R. J. (1973). Robert Bakewell (1725–1795), pioneer animal breeder, and his influence on Charles Darwin. *Folia Mendeliana* 58, 231–242.
- Xu, L., Bickhart, D. M., Cole, J. B., Schroeder, S. G., Song, J., Tassell, C. P., et al. (2014). Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol. Biol. Evol.* doi: 10.1093/molbev/msu333

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Orozco-terWengel, Barbato, Nicolazzi, Biscarini, Milanese, Davies, Williams, Stella, Ajmone-Marsan and Bruford. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.