npg

ORIGINAL ARTICLE

Multiple paternal origins of domestic cattle revealed by Y-specific interspersed multilocus microsatellites

L Pérez-Pardal[1,9], LJ Royo[1], A Beja-Pereira[2], S Chen[2], RJC Cantet[3], A Traoré[4], I Curik[5], J Sölkner[6], R Bozzi[7], I Fernández[1], I Álvarez[1], JP Gutiérrez[8], E Gómez[1], FA Ponce de León[9] and F Goyache[1]

[1]Área de Genética y Reproducción Animal, SERIDA, Camino de Rioseco, Gijón, Spain; [2]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, Vairão, Portugal; [3]Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires - CONICET, Buenos Aires, Argentina; [4]INERA, Ouagadougou, Burkina Faso; [5]Animal Science Department, University of Zagreb, Svetošimunska, Zagreb, Croatia; [6]Institute of Livestock Sciences, University of Natural Resources and Applied Life Sciences Vienna, Vienna, Austria; [7]Dipartimento Biotecnologie Agrarie, Sezione Scienze Animali, Università degli Studi di Firenze, Firenze, Italy; [8]Departamento de Producción Animal, Universidad Complutense de Madrid, Madrid, Spain and [9]Department of Animal Science, University of Minnesota, St Paul, MN, USA

In this study, we show how Y-specific interspersed multilocus microsatellites, which are loci that yield several amplified bands differing in size from the same male individual and PCR reaction, are a powerful source of information for tracing the history of cattle. Our results confirm the existence of three main groups of sires, which are separated by evolutionary time and clearly predate domestication. These three groups are consistent with the haplogroups previously identified by Götherström et al. (2005) using five Y-specific segregating sites: Y1 and Y2 in taurine (Bos taurus) cattle and Y3 in zebu (Bos indicus) cattle. The zebu cattle cluster clearly originates from a domestication process that was geographically and temporally separated from that of taurine clusters. Our analyses further suggest that: (i) introgression of wild sire genetic material into domesticated herds may have a significant role in the formation of modern cattle, including the formation of the Y1 haplogroup; (ii) a putative domestication event in Africa probably included local Y2-like wild sires; (iii) the West African zebu cattle Y-chromosome may have partially originated from an ancient introgression of humped cattle into Africa; and (iv) the high genetic similarity among Asian zebu sires is consistent with a single domestication process.
Heredity (2010) 105, 511–519; doi:10.1038/hdy.2010.30; published online 24 March 2010

Introduction

Cattle history, inferred from mitochondrial DNA (mtDNA) markers (Bradley et al., 1996; Loftus et al., 1994, 1999; Troy et al., 2001; Edwards et al., 2007a; Achilli et al., 2008), suggests that cattle (Bos taurus and Bos indicus) domestication involved at least two genetically distinct auroch (Bos primigenius) species in the Near East and great Indus Valley. Most European cattle resulted from the expansion of a small cattle population from the Near East after domestication, with one (T3) of the four major haplogroups (T, T1, T2 and T3) identified in the Near East predominating in the mainland continent. However, the history of European cattle may be more complex than what was previously thought. Recent studies have identified new mitochondrial haplogroups in European cattle (Achilli et al., 2008), and probable introgression events of local wild aurochs into domesticated B. taurus cattle (Beja-Pereira et al., 2006; Achilli et al., 2008). African cattle mtDNA lineages are mainly assigned to the T1 haplogroup, which is very rare in the

Middle East and Anatolia. This could suggest an independent domestication event, although the hypothesis of an African demographic expansion of a few T1-like cattle domesticated in the Near East is still arguable (Beja-Pereira et al., 2006; Bradley and Magee, 2006). Regarding zebu cattle, mtDNA sequences allowed the identification of two major haplogroups: I1 and I2 (Baig et al., 2005; Lai et al., 2006; Lei et al., 2006; Achilli et al., 2008; Chen et al., 2010). These two mtDNA lineages are well represented in India, suggesting that zebu domestication involved at least two different wild female populations (Baig et al., 2005) or, more likely, a single domestication event in the Indus Valley with a subsequent introgression process of wild (I2) females into proto-domesticated herds (Chen et al., 2010).

The DNA studies regarding cattle Y-chromosomes are comparatively rare (Verkaar et al., 2004) and have mainly focused on the assessment of male-mediated migration patterns and admixture between B. taurus and B. indicus (Hanotte et al., 2000; Anderung et al., 2007; Edwards et al., 2007b) or the assessment of differences in diversity (Ginja et al., 2009; Kantanen et al., 2009). Recent analyses have identified five polymorphic sites on the cattle Y-chromosome (Götherström et al., 2005), allowing the identification of three clusters (Y1, Y2 and Y3) in contemporary cattle, with Y1 being more frequent in

B. taurus from north-western Europe, Y2 being dominant in *B. taurus* found in southern Europe and Anatolian cattle, and Y3 being exclusive to *B. indicus*. The Y2 haplotypes would represent the taurine cattle domesticated in the Fertile Crescent, whereas the Y1 haplogroup would represent European aurochs. This would indicate a greater genetic influence of the primitive European aurochs on the formation of modern cattle breeds in Europe (Beja-Pereira *et al.*, 2006; Achilli *et al.*, 2008). However, ancient DNA does not generally support significant crossbreeding between local European wild cattle and domesticated cattle (Edwards *et al.*, 2007a; Bollongino *et al.*, 2008). In addition, medieval Scandinavian samples largely belonged to haplogroup Y2 and, overall, this scenario would not be consistent with a conserved signal from the time of domestication (Svensson and Götherström, 2008). Therefore, the strong geographical differentiation between the Y haplogroups found in modern cattle breeds in Europe requires further explanation.

Cattle Y-chromosome studies are also affected by a lack of powerful sources of information. There are limited numbers of informative segregating sites and polymorphic Y-specific microsatellites (Götherström *et al.*, 2005; Ginja *et al.*, 2009). Liu *et al.* (2003) reported the existence of several interspersed multilocus microsatellites (IMMs). These IMMs are loci that yield several amplified bands differing in size from the same male individual, using a single PCR reaction performed using a single primer pair. These loci are assumed to be interspersed and a result of the repetitive nature of the Y-chromosome. The male-specific region of the Y-chromosome is a mosaic of heterochromatic sequences and three classes of euchromatic sequences: X-transposed, X-degenerate and ampliconic (Skaletsky *et al.*, 2003). The absence of recombination with the X-chromosome causes rapid degeneration by mutation, deletion and transposon invasion that accumulate through time due to the larger number of cell divisions required to produce male gametes, and due to inefficient repair (for example, Muller's ratchet) and inefficient selection (Charlesworth *et al.*, 2005; Graves, 2006; Bachtrog and Andolfatto, 2006; Van Laere *et al.*, 2008). The repetitive nature of the Y-chromosome has been hypothesised to be a mechanism for maintaining Y-chromosome genes in a non-recombining environment and may arise by a number of mechanisms, including sexual antagonism, genomic conflict and hemizygous exposure (Vallender and Lahn, 2004; Murphy *et al.*, 2006). When a section is duplicated, it may result in a new locus that could generate new alleles at the locus site over time (Butler *et al.*, 2005). In the hemizygous state, the latter will be difficult to identify and, therefore, only its presence or absence can be identified. The new allele may be present in some individuals but absent in others, allowing each locus to be treated as a polymorphic site. Therefore, the polymorphism can be assessed as the presence/absence of a given band across individuals. However, the value of IMMs in population genetics studies has not been explored yet (Luo *et al.*, 2007). Here, we tested a set of IMMs to ascertain their usefulness for cattle phylogeny. The consistency of the information provided by the IMMs, along with the cattle history inferred from previous genetic analyses, will be discussed.

Materials and methods

Selection of IMMs
A total of five microsatellite loci, previously reported as IMMs by Liu *et al.* (2003), were tested in both taurine and zebu samples for male specificity, paternal compatibility and repeatability scoring to ascertain their performance in cattle Y-chromosome variability assessment (Pérez-Pardal *et al.*, 2009a). The IMM behaviour was consistent across subspecies. When amplification occurred using female DNA or identical band amplification results were not obtained with different PCR conditions, the IMM was eliminated and was not considered for further analyses (Figure 1 and Supplementary Figure S2). The PCR amplification conditions are described in Table 1. The description of the pedigrees used to test the markers is given in Supplementary Figure S2. Laboratory analyses were carried out using Automatic Sequencers ABI-310 and ABI-3130 (Applied Biosystems, Foster City, CA, USA). Size characterisation of PCR products was carried out using GenMapper software (Applied Biosystems).

Statistical analyses
Observed haplotypes were analysed as follows:

(1) Application of the Bayesian procedures implemented in the programme MrBayes 3.1. (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). Two MCMC runs, starting from different random trees, were completed; each with 20 000 000 generations and four chains. All sample points before reaching convergence were discarded as burn-in samples. The remaining samples were used to generate a majority rule consensus tree, in which the percentage of samples recovering any particular clade represented the clade's posterior probability (Huelsenbeck and Ronquist, 2001).
(2) Construction of biologically meaningful phylogenies of all genotyped samples using the program Network 4.5 (available at http://www.fluxus-engineering.com/). To avoid reticulation, a reduced median algorithm (Bandelt *et al.*, 1995) was used to generate a .rmf file, and the median joining network method (Bandelt *et al.*, 1999) was applied to this file. Following the authors recommendations (Forster *et al.*, 2000), the 'frequency >1' option was applied to discard singly occurring Y types. Same weights were assigned to each haplotype.
(3) Computation of correspondence analysis using Proc Corresp of the SAS/STAT package (SAS Institute, Cary, NC, USA).

Unbiased estimates of haplotype diversity and their variances ($V_{(h)}$) were calculated for the whole sample and for each identified haplogroup as mentioned in the study by Nei (1987) (formulas 8.5 and 8.13 therein). The s.e. value of h was calculated by taking the square root of $V_{(h)}$.

The program Arlequin 3.11 (Excoffier *et al.*, 2005) was used to calculate the analysis of molecular variance.

Divergence times were estimated using the ρ statistic (the average number of mutations from derived haplotypes to a haplotype designated as ancestral for the haplogroup; Forster *et al.*, 1996) using the program NETWORK 4.5. The average distance to the node of interest (ρ) was transformed to absolute time estimates by
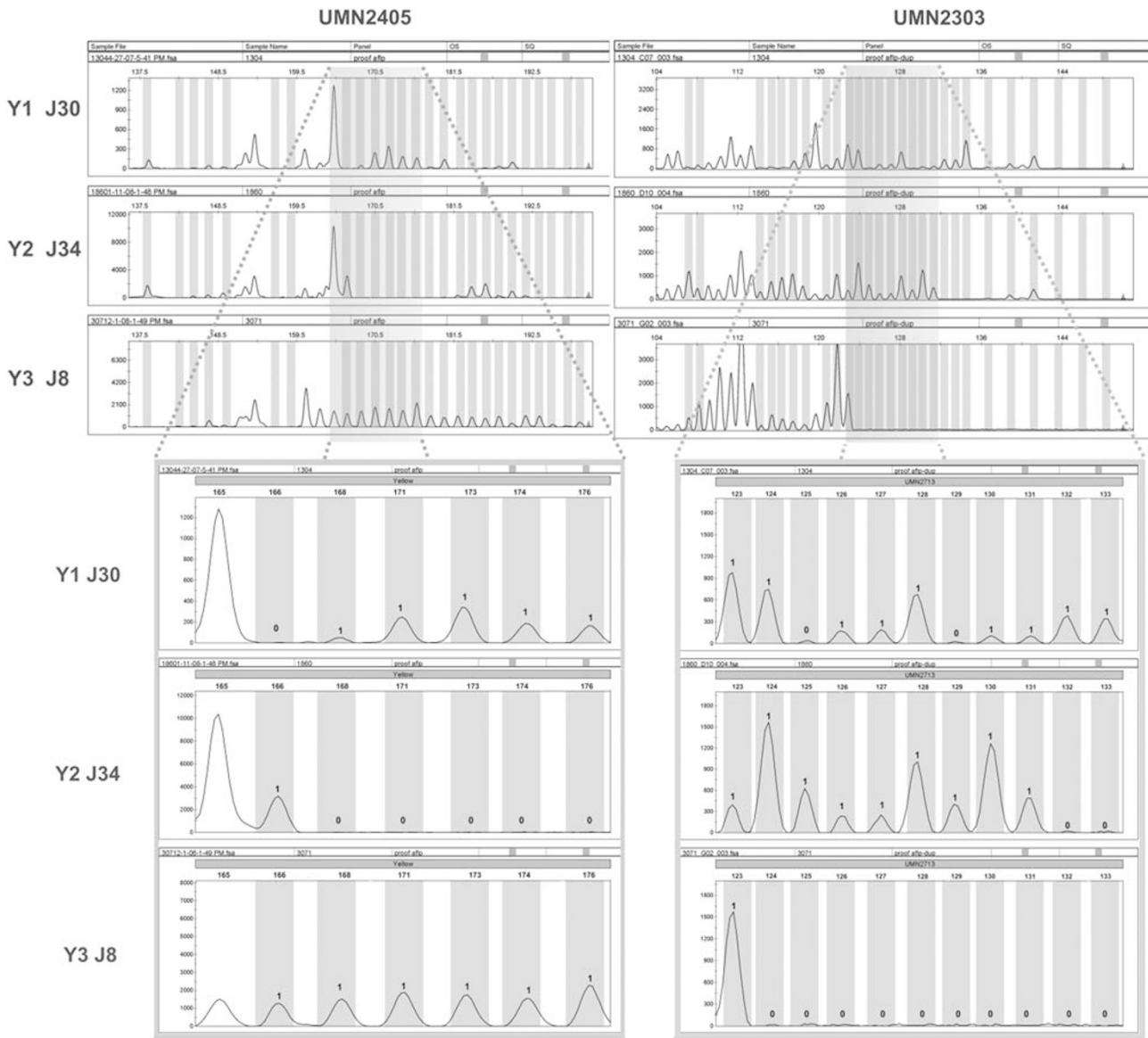
Figure 1 Electropherograms of cattle Y-specific interspersed multilocus microsatellites (IMMs) UMN2405 and UMN2303. We show haplotypes J30, J34 and J8 (see Supplementary Table 2) that belong to haplogroups Y1, Y2 and Y3, respectively. Grey bars represent the loci assessed as polymorphic across all samples used. Polymorphism is assessed as the presence/absence of a given band across individuals. Examples of genotyping that were carried out are shown by zooming in on parts of the electropherograms. Band sizes are shown at the top of the zoomed-in areas. The presence or absence of a given band in a sample is assessed as 1 (at the top of the peaks) or 0.

multiplication ($\rho \times$ years per mutation), assuming the phylogenies constructed using median-joining algorithms, as implemented in NETWORK. As a phylogeny-based statistic, ρ offers the advantage of being unbiased by demographic processes. The sampling error of ρ was approximated as $\sqrt{\rho/n}$, where n denotes the sample size. No published mutation rate for IMMs is available. As IMMs are defined as dominant markers, they should most probably be considered as slowly mutating markers. Therefore, we computed divergence times using the mutation rate (0.00018 per generation) reported by Forster *et al*. (2000) for slowly mutating Y-chromosome markers. The generation interval was fitted to 4.84 years. This value is the mean value of 15 different estimates reported by Gutiérrez *et al*. (2003) in eight breeds of cattle; most of them managed under very traditional conditions and not subject to sire selection programs.

The determination of ancestral haplotypes was performed by computing an internal number of square differences (d^2), computed as the average overall IMMs typed of $d_A^2 = \sum_{j=1}^{n} \frac{(j_a - j_b)^2}{n}$ where d_A^2 is the parameter for the IMM A, j_a and j_b are the lengths in repeat units of alleles a and b at IMM A, and n is the total number of bands at which IMM A was scored (here, 25 for IMM UMN2403 and 23 for UMN2303). Assuming a strict stepwise mutation model, the lower the value of d^2, the deeper the presence of the haplotype in the pedigree.

Results

We tested five IMMs reported by Liu *et al*. (2003) for the following: (i) amplification only with male DNA; (ii) polymorphism; (iii) paternal compatibility; and (iv)

Table 1 Description of IMMs tested

IMM	Primer sequences [5' to 3']	No. of bands		Size range in bp	Amplification with female DNA[a]	Paternal compatibility	Scoring repeatability[a]
		Total	Polymorphic				
UMN2405[b]	CCTGCCATCCATTGTGAAGA-F CTGCTTACCTGGTCAGGATT-R	30	25	138–190	No	Yes	Yes
UMN2303[b]	TACTTGCTTGAGACTTACTG-F TGTGAACACATCTGATTCTG-R	38	23	103–141	No[c]	Yes	Yes
UMN2713[d]	GTACCTACACTAATATGTTCA-F CCAAAGAAAGTTCAGGTACA-R	25	14	92–132	No	Yes	Yes[e]
UMN0920[d]	GTTGAGGACTCTTGCATCTG-F CACAGGCCTAGAAGATTGAG-R	22	Not tested	266–292	No	Not tested	No
UMN1201[d]	TGCTTCATCCTTCATTCCAC-F TTGTTGAGGACTCTTGCATC-R	28	Not tested	137–190	No	Not tested	No

Abbreviation: IMM, interspersed multilocus microsatellite.
Description of IMMs reported by Liu *et al*. (2003), which were tested for male specificity, paternal compatibility and scoring repeatability. Primer sequences, number of bands identified, number of polymorphic bands and corresponding base-pair size range are given. PCR was performed using a GenAmp PCR system 9700 (Applied Biosystems). The final reaction volume was 10 μl. The reaction mixture consisted of 50–100 ng of genomic DNA, 0.2 U of Taq polymerase (Biotools, Madrid, Spain), a final concentration of 0.2 μM of each primer (Sigma-Aldrich, Madrid, Spain), 200 μM of each dNTP (Biotools), and 2 mM MgCl$_2$. The PCR protocol included an initial step of 95 °C (3 min), followed by 30 cycles of 30 s at 95 °C for DNA denaturation, 30 s at 58 °C for primer annealing, and 30 s at 72 °C for primer extension. An additional final step of 72 °C for 90 min was included to avoid the presence of the *n*-1 band when using capillary electrophoresis. Samples from the Asturiana de los Valles family were genotyped three times for all markers using different PCR amplification conditions (primer annealing at 55, 58 and 60 °C) and included two different Taq polymerase enzymes (Biotools and TaqGold; Applied Biosystems) to verify the consistency of the results.
[a]Amplification with female DNA and scoring repeatability were tested in both *Bos taurus* and *Bos indicus* samples.
[b]Total number of bands and number of polymorphic bands assessed in the 453 samples analysed in the main study.
[c]IMM UMN2303 amplified a 96-bp-long amplicon using female DNA, which differs in size from those amplifications obtained from male DNA. Therefore, following Liu *et al*. (2003), the Y-specificity of IMM UMN2405 was accepted.
[d]Total number of bands and number of polymorphic bands found in the pedigrees described above (see also Supplementary Figure S2).
[e]IMM UMN2713 did not show consistent scoring repeatability across cattle subspecies. Therefore, it was not used for further analyses.

correct and repeatable scoring. Two of the tested IMMs (UMN2405 and UMN2303) fulfilled these requirements and were used further in this study. The IMM UMN2405 had a total of 30 bands, 25 of which were polymorphic. Of the 38 sites, 23 were polymorphic for UMN2303 (Table 1).

A total of 453 male individuals belonging to 52 cattle populations of Europe (351 samples), Africa (44 samples) and Asia (58 samples; Supplementary Table S1) were genotyped for the two selected IMMs. Altogether, they yielded a total of 146 haplotypes (haplotype diversity = 0.321 ± 0.005). Genotypes, haplotypes and their frequencies are provided in Supplementary Table S2.

Haplotypes were analysed by constructing evolutionary trees using Bayesian, network and correspondence analyses. All procedures allowed for the identification of three main haplogroups (Figure 2); two of these groups included only taurine samples, whereas the third was composed of only zebu samples. The consistency of these haplogroups with those previously identified by Göther-ström *et al*. (2005) using Y-specific single nucleotide polymorphisms was tested (see Supplementary Table S3). They were fully coincident with our haplogrouping. Therefore, the IMM-derived haplogroups will be referred to as Y1, Y2 (taurine) and Y3 (zebu). The Bayesian analysis provided a statistically high confidence value (0.76 between the taurine clusters and 1.0 among the taurine and zebu clusters). The observed haplogroups also showed high geographic consistency. The Y1 *B. taurus* haplogroup included those breeds located on the European Atlantic coastline and British Islands, and the Y2 *B. taurus* haplogroup included those breeds spread across continental Europe, the three European (Iberian, Italian and Balkan) peninsulas and Africa. The taurine Y-chromosomes are the only cattle Y-chromosome present in Europe and are the most frequent cattle chromosomes in Africa and Northern Asia (from the Eastern Caspian Sea to Mongolia and Japan). The Y3 *B. indicus* haplogroup was the only haplogroup present in India and was well represented in Western Africa.

Haplogroup Y1 had the highest haplotype diversity (0.482 ± 0.011), whereas both the zebu (Y3) and the main taurine (Y2) haplogroups had substantially lower haplotype diversities (0.276 ± 0.006 and 0.268 ± 0.014, respectively). The Y2 haplotypes identified in African cattle are not present on other continents. Interestingly, the only Japanese Kobe and Mongolian samples available to us tended to cluster separately within haplogroup Y2 (statistical confidence: 0.98; Supplementary Figure S1) with haplotypes identified in European breeds. Most Y1 haplotypes identified in Asian and African cattle are represented in Friesian cattle, suggesting that these haplotypes did not originate on these continents.

The average number of locus differences within and between haplogroups was calculated using the program Arlequin 3.11 (Excoffier *et al*., 2005). The higher mean pairwise differences within haplogroups for taurine were Y1 (5.08 ± 2.48) and Y2 (5.00 ± 2.44), whereas the difference was 2.58 ± 1.40 in zebu (Y3) haplotypes. The corrected mean pairwise differences between Y1 and Y2 with Y3 were 16.67 and 17.39, respectively. The difference between Y1 and Y2 was 12.06.
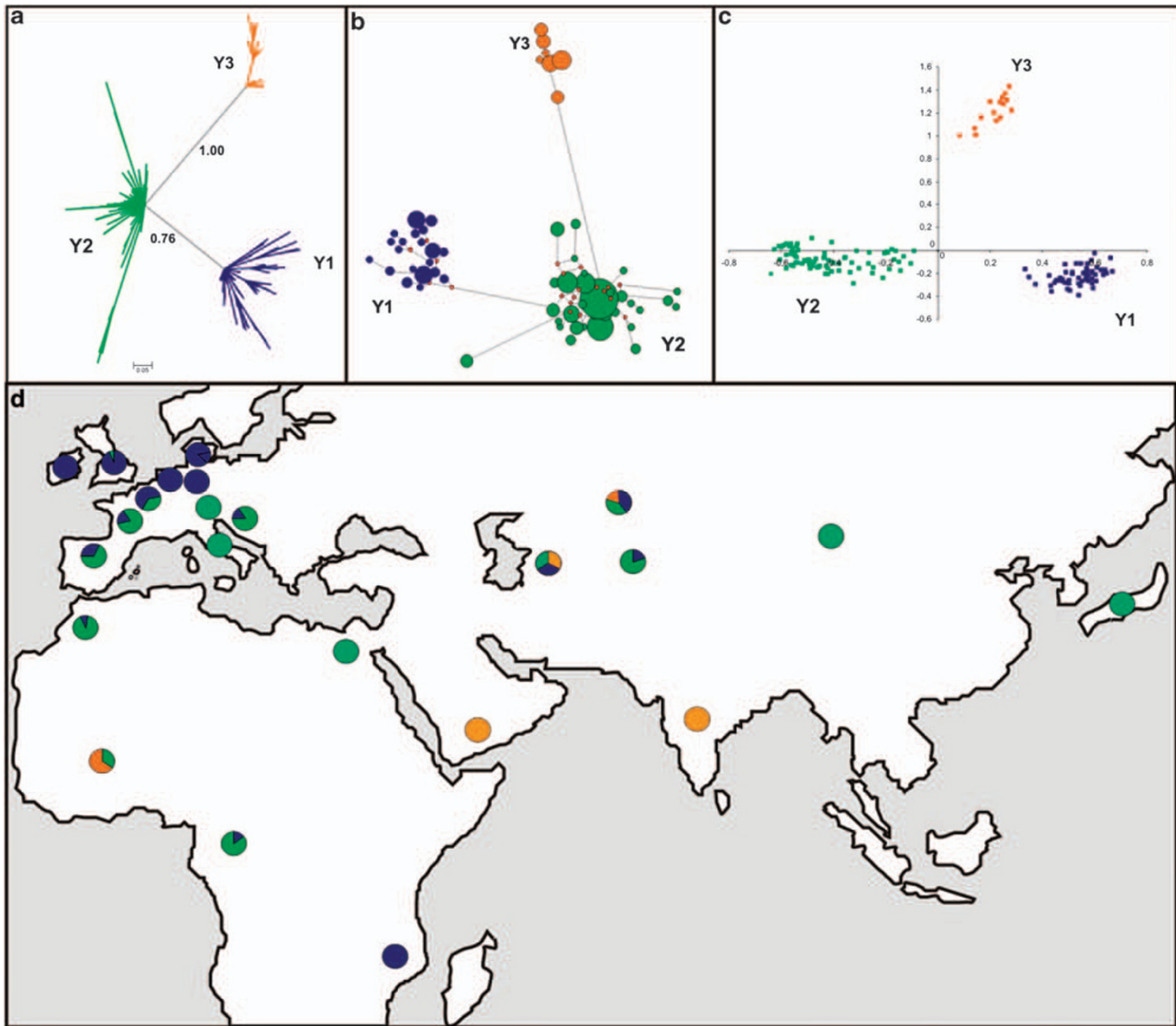
Figure 2 Graphical representation of genetic variability. (**a**) The phylogenetic tree constructed from the 146 identified haplotypes using the Bayesian procedures implemented in the program MrBayes 3.1. (**b**) A network tree constructed using the program Network 4.5. (**c**) The two dimensions were calculated on the polymorphism of the 146 Y-chromosome haplotypes identified through correspondence analysis. The three identified haplogroups were named Y1 (in blue), Y2 (in green) and Y3 (in orange). (**d**) Geographic map and frequency of haplogroups (Y1: blue; Y2: green; and Y3: orange) by country (see Supplementary Table S1). The numbers of samples and haplotypes included in each haplogroup were as follows: Y1: 114/54; Y2: 281/77; and Y3: 58/15.

Analysis of molecular variance (Excoffier *et al.*, 2005) was carried out at the haplogroup, continent and geographic area levels (Table 2). Analysis of molecular variance showed that intercontinental genetic differentiation was relatively weak (explaining approximately one-third of the variability), whereas variability was greater than 50% among the defined geographical areas. Breeds accounted for less than 30% of the variation.

Divergence times were estimated using ρ procedures (Forster *et al.*, 1996) assuming three different mutation rates (see Materials and methods) and a generation interval of 4.84 years (Table 3). The estimated divergence times between haplogroups, Y3 and Y1, varied from 131 761.0 ± 5378.0 to 658 780.5 ± 26 889.0 years before present (yBP). However, divergences between haplogroups Y1 and Y2 (varying from 77 981.0 ± 3802.8 to 389 890.5 ± 19 013.4 yBP) and between Y2 and Y3 (varying

from 75 292.0 ± 3802.8 to 376 446.0 ± 19 013.4 yBP) were substantially lower.

Discussion

To the best of our knowledge, we have, for the first time, used Y-specific IMMs to ascertain the history of an animal species from the male side. The benefits of using IMMs in cattle are clear; when compared with the available Y-specific microsatellites and single nucleotide polymorphisms (Götherström *et al.*, 2005; Ginja *et al.*, 2009; Kantanen *et al.*, 2009; Pérez-Pardal *et al.*, 2009a, b), IMMs have higher numbers of polymorphisms and higher resolution. Very recently, using classical Y-specific microsatellites we have been able to distinguished two different Y2 haplotypic subfamilies in taurine cattle; one of them restricted to the African continent (Pérez-Pardal

Table 2 Geographical structuring assessed using AMOVA

Type of grouping	Percentage of variance components			Total	ϕ Statistics		
	Among groups	Among populations within groups	Within populations	variance	ϕ_{ST}	ϕ_{SC}	ϕ_{CT}
Haplogroups (Y1, Y2 and Y3)	—	75.12 (2)[a]	24.88 (451)	9.460	0.751	—	—
Continents (Europe, Asia and Africa)	—	36.39 (2)	63.61 (451)	7.990	0.364	—	—
Geographical areas[b]	—	56.96 (5)	43.04 (448)	7.714	0.570	—	—
Breeds[c]	—	70.58 (47)	29.42 (397)	6.257	0.706	—	—
Breeds within continent	33.60 (2)	43.16 (45)	23.24 (397)	7.922	0.768	0.650	0.336
Breeds within geographic area	55.11 (5)	20.96 (42)	23.94 (397)	7.692	0.761	0.467	0.551

Abbreviation: AMOVA, analysis of molecular variance.
AMOVA subdivides the genetic diversity into hierarchical components and estimates the indices, ϕ, which are molecular equivalents of Wright's F statistics and can be interpreted as follows: ϕ_{ST}: relative divergence between populations; ϕ_{SC}: relative divergence between populations of the same group; ϕ_{CT}: relative divergence between groups. Note that the genetic variance was subdivided sequentially into one (haplogroups, continents, geographic areas or breeds) or two components (breeds within continents or geographic areas). The statistical significance of variance components and ϕ indices is evaluated by bootstrapping using 1000 replications. All estimates were statistically significant for $P < 0.0001$.
[a]Degrees of freedom are shown in parentheses.
[b]Geographical areas fitted were as follows (see Supplementary Table S1): Atlantic Europe (including the UK breeds, Dexter, Danish Red, Jersey, European and Holstein Friesian, Blanc-Bleu Belge, Normande, Asturian, and Tudanca cattle breeds); main European continent (including the rest of the European cattle breeds); Indian cattle (including Indian and Brahman samples); non-Indian Asia; Mediterranean Africa (including Moroccan and Egyptian samples); and Sub-Saharan Africa (including Lobi, N'Dama cattle and Landim samples).
[c]Populations with a sample size below three (Egypt, Landim, Mongolia and Kobe; see Supplementary Table S1) were excluded from analyses involving breeds.

Table 3 Divergence times between ancestral haplotypes

Ancestral haplotype	J12–J19	J12–J24	J19–J24
Haplogroup	Y3–Y2	Y3–Y1	Y2–Y1
P (s.d.)	14.00 (0.71)	24.50 (1.00)	14.50 (0.71)
Years (s.d.)	376 446.0 (19 013.4)	658 780.5 (26 889.0)	389 890.5 (19 013.4)

Abbreviation: IMM, interspersed multilocus microsatellite.
Divergence times between ancestral haplotypes estimated using ρ procedures as implemented in the program NETWORK 4.5, assuming that IMMs have a slow mutation rate (0.00018 per generation).

et al., 2009b). However, this set of markers was not powerful enough to correctly identify and differentiate haplogroups Y1 and Y2. This result underscores the need to identify new Y-specific markers to assess genetic relationships between different bovine Y-specific haplogroups (Pérez-Pardal et al., 2009b).

Moreover, the interest of this study extends beyond the bovine species, as IMMs have been identified in the Y-chromosomes of many other mammals, such as the family Felidae (Luo et al., 2007) and the genus Equus (Wallner et al., 2004). However, the use of IMMs has some constraints. First, their use is not straightforward: their selection, amplification, genotyping and scoring must be performed carefully. The second, and more important, problem is that their mutation model is currently unknown, which limits their use in evolutionary studies.

In this study, IMMs yield information about unanswered questions of cattle history from male legacies in light of previously available information: (i) what is the genetic influence of local aurochs on the formation of domesticated cattle?; (ii) do African cattle result from a local domestication event?; and (iii) are zebu cattle genetically uniform at the sire level?

Within Y2, an intriguing subcluster was identified. It includes cattle Y-chromosomes sampled in Mongolia, Japan, Northern Italy and Northern Atlantic Europe. Recently, Achilli et al. (2008) identified the new mtDNA haplogroup Q in Northern Italian and Korean cattle

samples, possibly resulting from the introgression of local wild aurochs. These are largely the same geographical areas in which our subcluster was found. Therefore, a likely explanation for the Y2-subcluster identified in this study is the introgression of wild sires into domesticated cattle. This introgression process would have involved poorly differentiated wild sire populations and would have happened in very distant geographical areas. This is consistent with mtDNA data that demonstrate the existence of a wild cattle population sharing many genetic features from Europe to North Asia (Achilli et al., 2008). Overall, the recruitment processes of wild cattle are likely to have occurred in some parts of Europe and Northern Asia, giving genetic signatures at both mtDNA and Y-chromosome levels.

On the basis of the strong geographic signal obtained from modern samples, Götherström et al. (2005) proposed the possible introgression of local aurochs into European domesticated cattle. However, more recent evidence obtained from ancient and medieval samples does not support this hypothesis (Bollongino et al., 2008; Svensson and Götherström, 2008), and the alternative model of a sudden introduction of domestic cattle into Central Europe without significant crossbreeding with local wild cattle remains unchallenged (Edwards et al., 2007a; Bollongino et al., 2008). Therefore, practical fixation of Y1 haplotypes in Northern Atlantic European cattle breeds would more likely be the result of drift

processes that occurred just before and at the moment of creation of modern cattle breeds approximately 200 yBP (Svensson and Götherström, 2008). However, this argument has an intuitive weakness because just after the formation of modern cattle breeds (approximately 100 yBP), the intense and sustained spreading of Y1-fixed cattle breeds (Friesian and British cattle) was facilitated by improved possibilities for long-distance trade and artificial insemination (approximately 50 yBP). The latter occurred all over the world, particularly across Europe. This dramatic and uncontrolled spreading process would have unavoidably erased differences among European taurine breeds at the Y-haplogroup level. It is noteworthy, that the Y1 samples identified in Africa and Northern Asia (Figure 2) are more likely to be the result of a recent introgression of European cattle rather than the expansion of a genetically heterogeneous sire population domesticated in the Fertile Crescent. Moreover, our results do not allow for the rejection of the hypothesis that introgression of local male aurochs partially influenced the formation of European cattle. This hypothesis is supported by two facts. The first is that the haplotypic diversity in the Y1 cluster is substantially higher than that observed for haplogroups Y2 and Y3, which are expected to be the result of domestication. This would suggest that the occurrence of a few local introgression events of Y1-related sires rather than a domestication process resulting in a much more intense bottleneck. The second fact is that ancient cattle management differences between northern and southern European cattle holders are supported by observed differences between northern and southern European cattle populations at the mtDNA level (Beja-Pereira et al., 2006), and evidence of co-evolution of cattle milk protein genes and human lactase in Northern Europe (Beja-Pereira et al., 2003). Our results support the idea that cattle management in Atlantic Europe might include the use of wild males with domesticated cows.

The question of the origins of both African taurine and zebu cattle has been the subject of intense debate and remains controversial. Our study confirms the existence of a Y2 haplotypic subfamily in African cattle (Pérez-Pardal et al., 2009b). Consistent with some archaeological findings (Wendorf and Schild, 1994), Bradley and Magee (2006) claim that African taurine cattle are not a simple subset of those domesticated in the Near East, due to the following three facts: (i) the mtDNA composition of cattle from these two areas is qualitatively different, with African samples belonging to the T1 haplogroup and showing a large proportion of unique haplotypes (Loftus et al., 1994; Bradley et al., 1996; Troy et al., 2001); (ii) there is a significant component of microsatellite variation within African cattle that seems to be indigenous in origin (Hanotte et al., 2002); and (iii) microsatellite variation also suggests that gene flow from the Near East into Northeast Africa was predominantly male-driven (MacHugh et al., 1997). Our results support this hypothesis because: (i) most haplotypes identified in African taurines are assigned to haplogroup Y2 and are not present in other continents; and (ii) the African Y2 samples show high haplotypic diversity (0.708 ± 0.069). At the mtDNA level, African cattle do not present the lack of variation associated with a severe foundation bottleneck, which would be necessary to fix a haplotype (T1) family after the domestication event occurred in the

Fertile Crescent (Bradley and Magee, 2006). This evidence can also be applied to our findings regarding the Y-chromosome.

The origin of African cattle is also linked to humped zebu cattle. Archaeological evidence suggests introgression of the humped zebu-like cattle into Africa approximately 3000 yBP (Payne, 1970). However, a major wave of B. indicus introgression in Africa may have started around the end of the seventh century AD in the Arab settlements along the East coast of Africa (Epstein, 1971), with subsequent spreading through zebu male-mediated interbreeding with African taurine females (Hanotte et al., 2000, 2002). Four of the six Y3 haplotypes identified in West African cattle tend to cluster together and are not present in Asian zebu sires (Supplementary Figure S1). This finding indicates a non-recent introgression of zebu cattle into Africa. Indeed, West African cattle share haplotypes with Asian zebu samples. However, West African cattle are often regarded as modern descendants of the founding African B. taurus cattle, which have been less affected by recent B. indicus introgression into the continent (Hanotte et al., 2002; Edwards et al., 2007b). Therefore, West African sanga cattle (the technical designation for the cross between taurine and zebu cattle) could be representative of a first ancient introgression of zebu cattle into Africa.

The scenario seen in our Asian zebu samples is quite different from that of cattle on the other continents. Most Y3 haplotypes observed in Asian zebu sires are present in the Indian populations, which leads to low haplotypic diversity (0.256 ± 0.017). At the maternal genetic level, a major component of zebu cattle diversity is retained by two distinct mitochondrial haplogroups (I1 and I2), suggesting two different domestication events or a single domestication process involving two different wild female populations (Baig et al., 2005; Lai et al., 2006; Chen et al., 2010). In this study, the sampling size for Asian zebu sires was limited. However, Indian samples were obtained in areas in which both I1 and I2 mtDNA clades are well represented (Chen et al., 2010). Despite this, haplotypic diversity is extremely low, suggesting high genetic homogeneity of Asian male zebus. This would support the hypothesis of a single domestication event for B. indicus. Recently, Chen et al. (2010) suggested that I1-like zebu cattle were domesticated in a single event in the Indus Valley (approximately 6000 yBP), with subsequent recruitment of mitochondrial I2 wild zebu females approximately 3500 yBP. Our results suggest that this introgression event of zebu wild females into proto-domesticated herds would have occurred with previously domesticated zebu sires.

To the best of our knowledge, our IMM data, for the first time, allowed the estimation of divergence times between cattle subspecies (B. taurus and B. indicus) on the male genetic path. Recent analyses of the complete B. taurus and B. indicus mitochondrial genome sequences estimated their divergence times to be 1.7–2.0 million years (Hiendleder et al., 2008). However, using a fragment of the HVI region of mtDNA, most divergence times between the two subspecies were estimated to be 200 000–1 000 000 yBP (Loftus et al., 1994 and others). This mtDNA fragment gives information that is comparable in amount to that given by our IMMs. Current estimates are highly consistent with the generally accepted range of divergence times (Loftus et al., 1994). However,

divergence times estimated between the Y1–Y2 and Y2–Y3 haplogroups are consistently lower than the estimated time between Y1–Y3. Overall, this would indicate a recent fragmentation of the wild populations that predated domestication, and a lack of differentiation between geographically distant wild sire populations.

In summary, we have demonstrated the usefulness of Y-specific IMMs to complement and add information to mtDNA-based studies regarding the history of the species. The divergent sire lineages, characterised here by Y-haplogroups, align with geographic areas and support previously suggested origins of cattle and directions of spread (Bradley et al., 1996; Troy et al., 2001; Baig et al., 2005; Lai et al., 2006; Edwards et al., 2007a; Achilli et al., 2008). In light of the above evidence, the most parsimonious interpretation of our data is that cattle have multiple paternal origins, possibly arising through multiple independent domestications. However, it is also possible that many of the paternal lineages identified in this study originated through introgression and not through separate domestication events.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

References

Achilli A, Olivieri A, Pellecchia M, Uboldi C, Colli L, Al-Zahery N et al. (2008). Mitochondrial genomes of extinct aurochs survive in domestic cattle. Curr Biol 18: R157–R158.

Anderung C, Hellborg L, Seddon J, Hanotte O, Göthersström A (2007). Investigation of X- and Y-specific single nucleotide polymorphisms in taurine (Bos taurus) and indicine (Bos indicus) cattle. Anim Genet 38: 595–600.

Bachtrog D, Andolfatto P (2006). Selection, recombination and demographic history in Drosophila miranda. Genetics 174: 2045–2059.

Baig M, Beja-Pereira A, Mohammad R, Kulkarni K, Farah S, Luikart G (2005). Phylogeography and origin of Indian domestic cattle. Curr Sci 89: 38–40.

Bandelt HJ, Forster P, Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16: 37–48.

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995). Mitochondrial portraits of human populations. Genetics 141: 743–753.

Bradley DG, MacHugh DE, Cunningham P, Loftus RT (1996). Mitochondrial diversity and the origins of African and European cattle. Proc Natl Acad Sci USA 93: 5131–5135.

Bradley DG, Magee DA (2006). Genetics and origins of domestic cattle. In: Zeder MA, Bradley DG, Emshwiller E, Smith BD (eds). Documenting domestication New genetic and archaeological paradigms. University of California Press: Berkeley-Los Angeles-London, pp 317–328.

Beja-Pereira A, Caramelli D, Lalueza C, Vernesi C, Ferrand N, Sampietro L et al. (2006). Genetic evidence for multiple origins of European cattle in Near-East, Africa, and Europe. Proc Natl Acad Sci USA 21: 8113–8118.

Beja-Pereira A, Luikart G, England PR, Bradley DG, Jann OC, Bertorelle G et al. (2003). Gene-culture coevolution between cattle milk protein genes and human lactase genes. Nat Genet 35: 311–313.

Bollongino R, Elsner J, Vigne J-D, Burger J (2008). Y-SNPs do not indicate hybridisation between European aurochs and domestic cattle. PLoS ONE 3, http://dx.doi.org/10.1371%2Fjournal.pone.0003418.

Butler JM, Decker AE, Kline MC, Vallone PM (2005). Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation. J Forensic Sci 50: 853–859.

Charlesworth D, Charlesworth B, Marais G (2005). Steps in the evolution of heteromorphic sex chromosomes. Heredity 95: 118–128.

Chen S, Lin B-Z, Baig M, Mitra B, Lopes RJ, Tarroso P et al. (2010). Zebu cattle are an exclusive legacy of the South Asia Neolithic. Mol Biol Evol 27: 1–6.

Edwards CJ, Baird JF, MacHugh DE (2007b). Taurine and zebu admixture in Near Eastern cattle: a comparison of mitochondrial, autosomal and Y-chromosomal data. Anim Genet 38: 520–524.

Edwards CJ, Bollongino R, Scheu A, Chamberlain A, Tresset A, Vigne J-D et al. (2007a). Mitochondrial DNA analysis shows a near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. Proc Biol Sci B 274: 1377–1385.

Epstein H (1971). The Origins of the Domestic Animals of Africa, vol. 1. Africana Publishing Corporation: London.

Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol Bioinf Online 1: 47–50.

Forster P, Harding R, Torroni A, Bandelt H-J (1996). Origin and evolution of native American mtDNA variation: a reappraisal. Am J Human Genet 59: 935–945.

Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C et al. (2000). A short tandem repeat-based phylogeny for the human Y chromosome. Am J Hum Genet 67: 182–196.

Ginja C, Telo da Gama L, Penedo MCT (2009). Y Chromosome haplotype analysis in Portuguese cattle breeds using SNPs and STRs. J Hered 100: 148–157.

Göthersström A, Anderung C, Hellborg L, Elburg R, Smith C, Bradley DG et al. (2005). Cattle domestication in the near East was followed by hybridization with aurochs bulls in Europe. Proc R Soc B 272: 2345–2350.

Graves JAM (2006). Sex chromosome specialization and degeneration in mammals. Cell 124: 901–914.

Gutiérrez JP, Altarriba J, Díaz C, Quintanilla R, Cañón J, Piedrafita J (2003). Pedigree analysis of eight Spanish beef cattle breeds. Genet Sel Evol 35: 43–63.

Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JE (2002). African pastoralism: genetic imprints of origins and migrations. Science 296: 336–339.

Hanotte O, Tawah CL, Bradley DG, Okomo M, Verjee Y, Ochieng J et al. (2000). Geographic distribution and frequency of a taurine Bos taurine and an indicine Bos indicus Y-specific

allele amongst sub-Saharan African cattle breeds. *Mol Ecol* **9**: 387–396.

Hiendleder S, Lewalski H, Janke A (2008). Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication. *Cytogenet Genome Res* **120**: 150–157.

Huelsenbeck JP, Ronquist F (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**: 754–755.

Kantanen J, Edwards CJ, Bradley DG, Viinalass H, Thessler S, Ivanova Z *et al.* (2009). Maternal and paternal genealogy of Eurasian taurine cattle (*Bos taurus*). *Heredity* **103**: 404–415.

Lai SJ, Liu YP, Liu YX, Li XW, Yao YG (2006). Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loop sequence variation. *Mol Phyl Evol* **38**: 146–154.

Lei CZ, Chen H, Zhang HC, Cai X, Liu RY, Luo LY *et al.* (2006). Origin and phylogeographical structure of Chinese cattle. *Anim Genet* **37**: 579–582.

Liu W-S, Beattie CW, Ponce de León FA (2003). Bovine Y chromosome microsatellite polymorphism. *Cytogenet Genome Res* **10**: 53–58.

Loftus R T, Ertugrul O, Harba AH, El-Barody MAA, MacHugh DE, Park SDE *et al.* (1999). A microsatellite survey of cattle from a centre of origin: the Near East. *Mol Ecol* **8**: 2015–2022.

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.

Luo S-J, Johnson WE, David VA, Menotti-Raymond M, Stanyon R, Cai QX *et al.* (2007). Development of Y chromosome intraespecific polymorphic markers in the Felidae. *J Hered* **98**: 400–413.

MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG (1997). Microsatellite DNA variation and the evolution, domestications and phylogeography of taurine and zebu cattle. *Genetics* **146**: 1071–1086.

Murphy WJ, Wilkerson AJP, Raudsepp T, Agarwala R, Schaffer AA, Stanyon R *et al.* (2006). Novel gene acquisition on carnivore Y chromosomes. *PLoS Genet* **2**: 353–363.

Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.

Payne WJA (1970). *Cattle Production in the Tropics*, 1st edn. Longman: London.

Pérez-Pardal L, Royo LJ, lvarez I, Ponce de León FA, Fernández I, Casais R *et al.* (2009a). Female segregation patterns of the putative Y-chromosome specific microsatellite markers INRA124 and INRA126 do not support their use for cattle population studies. *Anim Genet* **40**: 560–564.

Pérez-Pardal L, Royo LJ, Beja-Pereira A, Curik I, Traoré A, Fernández I *et al.* (2009b). Y-specific microsatellites reveal an African subfamily in taurine (*Bos taurus*) cattle. *Anim Genet*, doi: 10.1111/j.1365-2052.2009.01982.x.

Ronquist F, Huelsenbeck JP (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG *et al.* (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**: 825–837.

Svensson E, Götherström A 2008. Temporal fluctuations of Y-chromosomal variation in *Bos taurus*. *Biol Lett* **4**: 752–754.

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P *et al.* (2001). Genetic evidence for near-Eastern origins of European cattle. *Nature* **410**: 1088–1091.

Vallender EJ, Lahn BT (2004). How mammalian sex chromosomes acquired their peculiar gene content. *Bioessays* **26**: 159–169.

Van Laere AS, Coppieters W, Georges M (2008). Characterization of the bovine pseudoautosomal boundary: documenting the evolutionary history of mammalian sex chromosomes. *Genome Res* **18**: 1884–1895.

Verkaar ELC, Nijman IJ, Beeke M, Hanekamp E, Lenstra JA (2004). Maternal and paternal lineages in cross-breeding bovine species. Has wisent a hybrid origin? *Mol Biol Evol* **21**: 1165–1170.

Wallner B, Piumi F, Brem G, Müller B, Achmann R (2004). Isolation of Y chromosome-specific microsatellites in the horse and cross-species amplification in the genus *Equus*. *J Hered* **95**: 158–164.

Wendorf F, Schild R (1994). Are the early Holocene cattle in the Eastern Saharan domestic or wild? *Evol Anthropol* **3**: 118–128.

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