



Y-specific microsatellites reveal an African subfamily in taurine (*Bos taurus*) cattle

L. Pérez-Pardal^{*,†}, L. J. Royo^{*}, A. Beja-Pereira[†], I. Curik[‡], A. Traoré[§], I. Fernández^{*}, J. Sölkner[¶], J. Alonso^{**}, I. Álvarez^{*}, R. Bozzi^{††}, S. Chen[†], F. A. Ponce de León^{‡‡} and F. Goyache^{*}

^{*}Área de Genética y Reproducción Animal, SERIDA, Camino de los Claveles 604 33203-Gijón, Spain. [†]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal. [‡]Animal Science Department, University of Zagreb, Svetošimunska, Zagreb, Croatia. [§]INERA, 04 BP 8645 Ouagadougou 04, Burkina Faso. [¶]BOKU – University of Natural Resources and Applied Life Sciences Vienna, Vienna, Austria. ^{**}Centro de Inteligencia Artificial, Universidad de Oviedo, Campus de Viseques, 33271-Gijón, Spain. ^{††}Dipartimento di Scienze Zootecniche, Università degli Studi di Firenze, Via delle Cascine 5, 50144 Firenze, Italy. ^{‡‡}Department of Animal Science, University of Minnesota, St. Paul, MN, 55108, USA

Summary

Five cattle Y-specific microsatellites, totalling six loci, were selected from a set of 44 markers and genotyped on 608 *Bos taurus* males belonging to 45 cattle populations from Europe and Africa. A total of 38 haplotypes were identified. Haplogroups (Y1 and Y2) previously defined using single nucleotide polymorphisms did not share haplotypes. Nine of the 27 Y2-haplotypes were only present in African cattle. Network and correspondence analyses showed that this African-specific subfamily clustered separately from the main Y2-subfamily and the Y1 haplotypes. Within-breed genetic variability was generally low, with most breeds (78%) showing haplotypes belonging to a single haplogroup. AMOVA analysis showed that partitioning of genetic variation among breeds can be mainly explained by their geographical and haplogroup assignment. Between-breed genetic variability summarized via Principal Component Analysis allowed the identification of three principal components explaining 94.2% of the available information. Projection of principal components on geographical maps illustrated that cattle populations located in mainland Europe, the three European Peninsulas and Mediterranean Africa presented similar genetic variation, whereas those breeds from Atlantic Europe and British Islands (mainly carrying Y1 haplotypes) and those from Sub-Saharan Africa (belonging to Y2-haplogroup) showed genetic variation of a different origin. Our study confirmed the existence of two large Y-chromosome lineages (Y1 and Y2) in taurine cattle. However, Y-specific microsatellites increased analytical resolution and allowed at least two different Y2-haplotypic subfamilies to be distinguished, one of them restricted to the African continent.

Keywords cattle, domestication, phylogeny, Y-chromosome, Y-specific markers.

Introduction

Putative microsatellites mapping to the male-specific Y-chromosome (MSY) region of cattle are available (Bishop *et al.* 1994; Vaiman *et al.* 1994; Mezzelani *et al.* 1995; Kappes *et al.* 1997; Liu *et al.* 2003). Y-specific microsatellites have been used to trace cattle domestication and differentiation among bovid species (Edwards *et al.* 2000,

2007a). Some of these markers were also used to distinguish between taurine (*Bos Taurus*) and zebu (*B. indicus*) patriline and to infer introgression events from zebu into taurine populations (Bradley *et al.* 1994; Edwards *et al.* 2000, 2007a; Giovambattista *et al.* 2000; Hanotte *et al.* 2000; Li *et al.* 2007). Götherström *et al.* (2005) identified two Y-specific single nucleotide polymorphisms (SNPs) allowing two major clusters to be distinguished in taurine (*B. taurus*) cattle (Y1 and Y2). The frequencies of the identified SNPs in modern samples revealed that haplogroup Y1 can primarily be found in extant northern European cattle breeds and haplogroup Y2 is mainly distributed in the Central and Southern European cattle breeds. This strong geographical signal could be explained by a significant

Address for correspondence

Goyache, Área de Genética y Reproducción Animal, SERIDA, Camino de los Claveles 604 33203-Gijón, Spain.

E-mail: foyache@serida.org

Accepted for publication 10 September 2009

influence of male aurochs on domesticated populations coming from the Near East during the formation of the present European cattle stock. This would challenge the common view on early domestication and Neolithic stock-rearing. However, Bollongino *et al.* (2008) and Svensson & Götherström (2008), using ancient and mediaeval samples, did not support the hypothesis that these markers distinguish European aurochs from domesticated cattle.

Y-specific microsatellites and SNPs have been jointly used to ascertain genetic history of human populations (Zerjal *et al.* 2002). SNPs provide an opportunity to understand paternal population origins, relationships and dispersals with greater phylogenetic and geographical resolution and less terminological ambiguity than was hitherto possible. The inclusion of microsatellite data allows the assessment of patterns of migration, admixture and ancestry, as well as to identify additional micro-evolutionary processes associated with population structure *sensu lato* (Zegura *et al.* 2004). Recently, extended Y-chromosome haplotypes that combine information from microsatellites and SNPs have been used to study genetic relationships among Portuguese (Ginja *et al.* 2009), Ethiopian (Li *et al.* 2007) and Eurasian (Kantanen *et al.* 2009) cattle populations. However, samples used are limited to these geographical areas.

The aim of this work was to investigate the haplotype structure and diversity in taurine cattle breeds from Europe and Africa using a combination of already described SNPs and microsatellites tested to be specific to the MSY region. The genetic relationships between breeds will also be analysed.

Materials and methods

Samples

DNA samples from 608 *B. taurus* males belonging to 45 cattle populations from Europe and Africa were analysed (Table 1). European sampling includes major breeds located in: (i) Atlantic Europe (12 breeds and 198 samples) including British, English Channel (including Normande) and European Friesian breeds; (ii) Continental Europe, including French (except Normande), German (except for European Friesian samples) and Austrian cattle breeds (13 breeds and 181 samples); (iii) Iberian Peninsula (16 breeds and 266 samples); (iv) Italian Peninsula (three breeds and 45 samples); and Balkan Peninsula (two breeds and 17 samples). African samples included two breeds from Mediterranean Africa (obtained in Morocco) totalling 10 samples and two Sub-Saharan Africa breeds (Lobi from Burkina Faso and N'Dama originated in Guinea) totalling 17 samples. Further description of sampling is given in Table 1.

SNP analyses

Three polymorphisms previously described by Götherström *et al.* (2005) were analysed as follows: (i) polymorphism

DDX3Y1 (DBY), which differentiates zebu and taurine Y-specific haplogroups, was analysed on the African samples to ensure that further determinations were carried out on *B. taurus* Y-chromosomes; (ii) polymorphism *zfy10* was analysed on all samples to distinguish between haplogroups Y1 and Y2; and (iii) polymorphism *UTY19* was analysed on the African Y2 samples to confirm the haplogroup assignment. For SNP *UTY19* (C/A), PCR conditions were as follows: the final reaction volume was 10 µl; the reaction mixture consisted on 50–100 ng of genomic DNA, 0.2 U of Taq polymerase (Biotools), 0.25 µM of each primer (Sigma-Aldrich), 200 µM of each dNTP (Biotools) and 2 mM MgCl₂. The PCR protocol included an initial step of 95 °C (3 min), followed by 40 cycles of 30 s at 95 °C for DNA denaturation, 30 s for primer annealing at 54 °C and 30 s at 72 °C for primer extension. PCR product was purified with the exoSAP-IT[®] protocol (USB) and sequenced in both strands by using the BigDye[®] Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) following the manufacturer's recommendations, on a ABI3130 sequence analyser (Applied Biosystems). A DX3Y1 C/T polymorphism (AY928811:g.437) differentiates taurine and zebu Y-chromosome, and an UTY G/T polymorphism (AY936543:g.423) differentiates Y1 and Y2 haplogroup. For the *ZFY10* indel, the PCR protocol was as follows: the final reaction volume was 10 µl; the reaction mixture consisted of 50–100 ng of genomic DNA, 0.2 units of Taq polymerase (Biotools), 0.25 µM of each primer (Sigma-Aldrich), 200 µM of each dNTP (Biotools) and 2 mM MgCl₂. 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 for primer annealing at 55 °C and 30 s at 72 °C for primer extension. An additional final step of 72 °C for 90 min was included to avoid n-1 band presence when using capillary electrophoresis (Table 2). Genotyping was carried out by direct electrophoreses on an ABI3130 sequencer analyser (Applied Biosystems) based on its difference in size: 285–287 nt for Y1 and Y2 respectively.

Microsatellite genotyping

A total of 43 microsatellites previously reported to be located on the bovine MSY region by Matthews *et al.* (1991), Bishop *et al.* (1994), Vaiman *et al.* (1994), Mezzelani *et al.* (1995), Kappes *et al.* (1997), Liu *et al.* (2003), and microsatellite *SRYM18* described on the ovine Y-chromosome by Meadows *et al.* (2006), were tested for male specificity, paternal compatibility and repeatability scoring following the protocol described in Pérez-Pardal *et al.* (2009; see Table S1). After this selection process only five microsatellites were found to be suitable for our study: *INRA189*, *UMN0103*, *UMN0307*, *BM861* and *BYM1*. Microsatellite *UMN0103* showed a bimodal allele-frequency distribution that could be scored as two different loci

Table 1 Distribution of the 608 male individuals analysed per breed and country. Each breed is assigned to a given geographical area: (i) Atlantic Europe; (ii) Continental Europe; (iii) Balkan Peninsula; (iv) Iberian Peninsula; (v) Italian Peninsula; (vi) Mediterranean Africa; and (vii) Sub-Saharan Africa. Assignment of samples to each identified haplogroup is also given (number of Y2 samples belonging to the African Y2 haplotypes are in brackets). Expected heterozygosity (He), haplotype diversity corrected for sample size (*h*) and standard errors (SE) are also given.

Breed	Country	Sample size	Geographical areas	Y1	Y2	He	<i>n</i>	<i>h</i>	SE(<i>h</i>)
Alistana	Spain	9	Iberian Peninsula		9	0	1	0.125	0.022
Angus	UK	13	Atlantic Europe	13		0.083	2	0.167	0.030
Asturiana de la Montaña	Spain	19	Iberian Peninsula	18	1	0	1	0.056	0.002
Asturiana de los Valles	Spain	38	Iberian Peninsula	32	6	0.321	3	0.081	0.003
Blanc-Bleu Belge	Belgium	15	Atlantic Europe	15		0.116	2	0.143	0.021
Berrenda	Spain	4	Iberian Peninsula		5	0.213	3	0.750	0.344
Betizu	Spain	10	Iberian Peninsula		10	0	1	0.111	0.016
Blonde d'Aquitaine	France	5	Continental Europe		5	0.133	3	0.750	0.344
Brown Swiss	Switzerland	14	Continental Europe		14	0	1	0.077	0.004
Bruna dels Pirineus	Spain	11	Iberian Peninsula		11	0.066	2	0.200	0.043
Bovin d'Oulmes	Morocco	5	Mediterranean Africa		5	0.053	2	0.500	0.219
Bovin de Tidili	Morocco	5	Mediterranean Africa		5 (1)	0.347	2	0.500	0.219
Charolais	France	5	Continental Europe		5	0	1	0.250	0.094
Chianina	Italy	20	Italian Peninsula		20	0	1	0.053	0.002
Croatian Simmental	Croatia	13	Balkan Peninsula	3	10	0.237	2	0.167	0.030
Danish Red	Denmark	10	Atlantic Europe	9	1	0.120	2	0.222	0.053
Dexter	Ireland	4	Atlantic Europe	4		0	1	0.333	0.156
Fleckvieh	Switzerland	4	Continental Europe		4	0	1	0.333	0.156
European Friesian	Germany ¹	8	Atlantic Europe	8		0.036	2	0.286	0.084
Galloway	UK	2	Atlantic Europe		2	0.083	2	2.000	1.937
Hereford	UK	8	Atlantic Europe	7	1	0.146	2	0.286	0.084
Holstein	The Netherlands	37	Atlantic Europe	37		0	1	0.028	0.006
Istrian	Croatia	4	Balkan Peninsula		4	0	1	0.333	0.156
Jersey	The Netherlands	5	Atlantic Europe		5	0	1	0.250	0.094
Toro de Lidia	Spain	65	Iberian Peninsula		65	0.084	3	0.047	0.002
Limousin	France	4	Continental Europe		4	0.063	2	0.667	0.349
Lobi	Burkina Faso	8	Sub-Saharan Africa		8 (8)	0.307	1	0.143	0.030
Maremmiana	Italy	19	Italian Peninsula		19	0	1	0.056	0.002
Morucha	Spain	17	Iberian Peninsula		17	0.219	3	0.188	0.031
N'Dama	Guinea ²	9	Sub-Saharan Africa	2	7 (7)	0.300	3	0.375	0.110
Normande	France	39	Atlantic Europe	39		0	1	0.026	0.006
Pajuna	Spain	4	Iberian Peninsula	1	3	0.354	3	1.000	0.541
Parda de Montaña	Spain	26	Iberian Peninsula	1	25	0.049	2	0.080	0.003
Parthenais	France	15	Continental Europe	4	11	0.261	2	0.143	0.021
Piemontese	Italy	6	Italian Peninsula		6	0.185	3	0.600	0.240
Pinzgauer	Austria	9	Continental Europe		9	0	1	0.125	0.022
Pirenaica	Spain	10	Iberian Peninsula		10	0.053	2	0.222	0.053
Retinta	Spain	6	Iberian Peninsula		6	0.148	2	0.400	0.151
Rubia Gallega	Spain	31	Iberian Peninsula		31	0	1	0.033	0.005
Salers	France	15	Continental Europe	1	14	0.083	2	0.143	0.021
Sayaguesa	Spain	4	Iberian Peninsula		4	0.125	3	1.000	0.541
Serrana Negra	Spain	3	Iberian Peninsula		3	0.148	2	1.000	0.663
Simmenthal	Germany	4	Continental Europe		4	0	1	0.333	0.156
Tarentaise	France	18	Continental Europe		18	0	1	0.059	0.001
Tudanca	Spain	8	Iberian Peninsula	8		0	1	0.143	0.030
Tyrolean Grey	Austria	19	Continental Europe		19	0.031	2	0.111	0.011
Totals				202	406 (16)		38	0.063	0.000

¹These samples were mainly from German sires. However, sires from Sweden and The Netherlands are included in this group.

²N'Dama samples, obtained in Congo from herds managed by the Compagnie JVL, have been assigned to Guinea because of the historical origin of the animals forming the sampled herds.

Table 2 Description of the Y-specific microsatellites and SNPs typed in this work. The microsatellites *INRA189*, *UMN0103*, *UMN0307*, *BM861* and *BYM1* were analysed in all samples. Two multiplex PCRs reactions (PCR 1: *INRA189* and *UMN0103* and PCR 2: *UMN0307*, *BM861* and *BYM1*) were carried out. Polymorphism *ZFY10* was analysed on all samples to distinguish between haplogroups Y1 and Y2. SNP *UTY19*, which differentiates zebu and taurine Y-specific haplogroups, was analysed on the African samples to ensure that further determinations are carried out on *Bos taurus* Y-chromosomes.

Microsatellite	Repeat pattern	Primer sequence	Primer concentration (μM)	Temperature ($^{\circ}\text{C}$)	Labelling	No. of alleles
INRA189	(TG) _n	TTTTGTTTCCCGTGCTGAG-F	0.25	58	VIC	8
		GAACCTCGTCTCCTTGTAGCC-R	0.25		No	
UMN0103	(CA) _n	ACACAGAGTATTCACCTGAG-F	0.25	58	FAM	9
		ATTACCTGGGTCAAAGCAC-R	0.25		No	
UMN0307	(CA) _n	GATACAGCTGAGTGACTAAC-F	0.18	58	PET	5
		GTGCAGACATCTGAGCTGTG-R	0.18		No	
BYM-1	(AC) _n	CCTTGTTTGAGCTTGACCACT-F	0.18	58	FAM	3
		TTGCAGGCACAGAAACGGA-R	0.18		No	
BM861	(GT) _n C(TG) _n	TTGAGCCACCTGGAAAGC-F	0.18	58	FAM	2
		CAAGCGGTTGGTTCAGATG-R	0.18		No	
DDX3Y1	–	GGAAACACAGAATCTTAATCTC-F	0.25	54	No	2
		AGTATGTCATTAGCCCATATC-R	0.25		No	
UTY 19	–	GGTAGGAAACAACCTTCTTTCTG-F	0.25	54	No	2
		GAACGTTCAAAGTTGTTTAC-R	0.25		No	
ZFY10	–	CCAAAATGGTTGAGCTTTATGA-F	0.25	55	FAM	2
		GGAGCATAAGTGATCCAATGAA-R	0.25		No	

(Balaesque *et al.* 2006). Therefore the five selected microsatellites included six different loci.

The five selected microsatellites were analysed on all samples in two multiplex PCRs (PCR 1: *INRA189* and *UMN0103* and PCR 2: *UMN0307*, *BM861* and *BYM1*) performed in 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 units of Taq polymerase (Biotools), 200 μM of each dNTP (Biotools) and 2 mM MgCl_2 . Concentration primers (Sigma-Aldrich) and fluorescent dye labels are given in Table 2. The PCR protocol included an initial step of 95 $^{\circ}\text{C}$ (3 min), followed by 30 cycles of 30 s at 95 $^{\circ}\text{C}$ for DNA denaturation, 30 s for primer annealing at 58 $^{\circ}\text{C}$ and 30 s at 72 $^{\circ}\text{C}$ for primer extension. An additional final step of 72 $^{\circ}\text{C}$ for 90 min was included to avoid n-1 band presence when using capillary electrophoresis. PCR fragments were separated by capillary electrophoresis on an ABI 310 instrument (Applied Biosystems) according to the manufacturer's recommendations. Allele sizes were determined with Genemapper (Applied Biosystems).

Terminology

The terms 'haplogroup' and 'haplotype' have various, overlapping definitions in the literature. Throughout the article, we will use the terminology proposed by de Knijff (2000) and recommended by the Y-Chromosome Consortium (2002) in which 'haplogroup' refers to MSY lineages defined by binary polymorphisms and the term 'haplotype' is

reserved for all subfamilies of haplogroups that are defined by variation at microsatellites on the MSY region.

The two different loci shown by microsatellite *UMN0103* will be analysed using the Mathias's prior (Balaesque *et al.* 2006). Mathias *et al.* (1994) assigned the co-amplified alleles to each locus according to their sizes, the shortest allele being assigned to locus 1, the longest to locus 2.

Statistical analyses

Microsatellite alleles were combined into haplotypes. Observed haplotypes were analysed as follows: (i) analysis of correspondence was performed using the Proc CORRESP of the SAS/STAT package (SAS Institute Inc.); and (ii) a median-joining network connecting different haplotypes was constructed using the program NETWORK 4.5.2 [available at <http://www.fluxus-engineering.com/> (Bandelt *et al.* 1999)]. 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.

An unbiased estimate of haplotype diversity, h , and its variance, $V(h)$, were calculated according to the method of Nei (1987, formulas 8.5 and 8.13 therein). The standard error of h , $SE(h)$, was calculated by taking the square root of $V(h)$. Single-locus gene-diversity values were calculated in the same way. Within-breed expected heterozygosity adjusted for sampling size was also computed according to Nei (1987). The between-breeds genetic identity matrix was computed, according to Nei (1987), as the average across loci of the term to

$$\sum_{ij} x_{ij} y_{ij}$$

where x_{ij} and y_{ij} are the frequencies of the i th allele at the j th locus within the populations x and y using the program MOLKIN (Gutiérrez *et al.* 2005). Information provided by this matrix was summarized by conducting principal component analyses (PCA) on the complement of this matrix. The principal component scores were used to construct interpolation maps drawn using the SPATIAL ANALYST EXTENSION OF ARCVIEW, available at: <http://www.esri.com/software/arcview/>. The Inverse Distance Weighted (IDW) option with a power of two was selected for the interpolation of the surface. IDW assumes that each input point has a local influence that diminishes with distance. The area of sampling of each breed was used as geographical coordinates, and the six nearest neighbours were used for the calculation. Interpolation surfaces were divided into eight equal classes.

AMOVA analyses were computed using the program ARLEQUIN 3.11 (Excoffier *et al.* 2005). Analyses were carried out fitting one single (haplogroup, haplofamily, breed, continent or geographical area) or two hierarchical levels (breed within continent or breed within geographical area).

Results

The number of alleles per polymorphic microsatellite varied from 2 (*BM861*) to 8 (*INRA189*), revealing a total of 32 alleles across the six loci used. Overall gene diversity was 0.429. Altogether, the identified alleles defined 38 haplotypes (overall haplotype diversity of 0.063 ± 0.0002). Haplotype descriptions and their distribution across breeds are given in Table S2. Samples were assigned to the Y-chromosome haplogroups defined by Götherström *et al.* (2005) according to the SNPs typed. *UTY19* polymorphism confirmed the Y2-haplogroup assignment of the African samples. A total of 11 haplotypes belonged to haplogroup Y1 (named from Y1_1 to Y1_11) whereas 27 haplotypes (from Y2_1 to Y2_27) belonged to haplogroup Y2.

Haplogroups did not share haplotypes. The most common Y1-haplotype (Y1_6) was identified in 106 samples belonging to seven different breeds. The most common Y2-haplotype (Y2_19) was identified in 269 samples across 29 breeds. Three Y1-haplotypes and eleven Y2-haplotypes were unique. Haplotype Y1_6 is basically fixed in the Normande, Holstein and European Friesian breeds. Haplotype Y2_19 was widely distributed across breeds, being dominant in Central and Southern European and Moroccan samples. A group of nine Y2-haplotypes (Y2_1, Y2_3, Y2_4, Y2_5, Y2_7, Y2_9, Y2_10, Y2_11 and Y2_27) were only present in Sub-Saharan cattle but also in one Moroccan sample (haplotype Y2_1).

Figure 1 summarizes the genetic variability contained by the 38 haplotypes identified. Correspondence analysis separated the main group of Y2-haplotypes from the Y1-haplotypes on Dimension 1 (X-axis). Interestingly

enough, the subfamily formed by the nine African specific Y2-haplotypes were also separated from the main Y2-subfamily on the X-axis; differentiation between this second Y2-subfamily and the Y1 haplotypes was assessed on Dimension 2 (Y-axis). The existence of 2 subfamilies within the Y2-haplogroup was confirmed via NETWORK analysis. The Y2-subfamily identified on African sire Y-chromosomes clustered with the Y1-haplotypes.

Within-breed genetic variability was generally low (Table 1). Most breeds (78%) showed haplotypes belonging to a single haplogroup. Seventeen (37%) of the analysed breeds showed a single haplotype. Within-breed expected heterozygosity varied from 0 to 0.354 (Pajuna breed). Within-breed haplotype diversity showed more variation, ranging from 0.026 ± 0.006 in the Normande breed to 1.0 in some endangered local breeds such as Sayaguesa, Serrana Negra and Pajuna breeds. In general, breeds that are considered closely related, such as those belonging to Brown Swiss type (Brown Swiss, Tyrolean Grey, Parda de Montaña and Bruna dels Pirineus) or the classical European Friesian and the modern Holstein cattle, have consistently the same haplotype composition, the most frequent being Y2_19 and Y1_6 respectively. Between-breed genetic relationships were summarized via PCA performed on the complement of the between-breeds genetic identity matrix (Table S3). Three principal components (PC) with eigenvalue >1 , explaining 81.8%, 10.2% and 2.2% of the total variability, were retained (Table S2). PC1 included mainly the Central European, Southern European and Moroccan breeds; PC2 mainly included Atlantic European breeds; PC3 was almost exclusively formed by Sub-Saharan Lobi and N'Dama breeds. This pattern was consistent with the variability previously shown by the identified haplotypes. Projection of PCs on a geographical map is shown in Fig. 2. The three PCs gave very similar images: cattle populations located in mainland Europe, the three European Peninsulas and Mediterranean Africa presented similar genetic variation whereas those breeds from Atlantic Europe and British Islands (mainly carrying Y1 haplotypes) and those from Sub-Saharan Africa (belonging to Y2-haplogroup) showed genetic variation of different origin.

Genetic identity between Sub-Saharan Africa populations and most other populations (regardless they are formed by 'European' Y2 samples) was below 0.30. This is only comparable with the genetic identity assessed between the Atlantic European sample and the other geographical cattle populations, ranging from 0.322 to 0.426 (Table 3). Samples obtained from breeds of the mainland European continent, the three European Peninsulas and Mediterranean Africa have between-population genetic identities ranging from 0.664 to 0.873. Atlantic European and Sub-Saharan samples have poor genetic identity (0.315).

This general scenario was confirmed by an AMOVA analysis (Table 4). The 'Continent' only accounted for roughly 20% of the variation and 'Geographical Area' explained less than 50% of the variation. When 'Breed' was included in

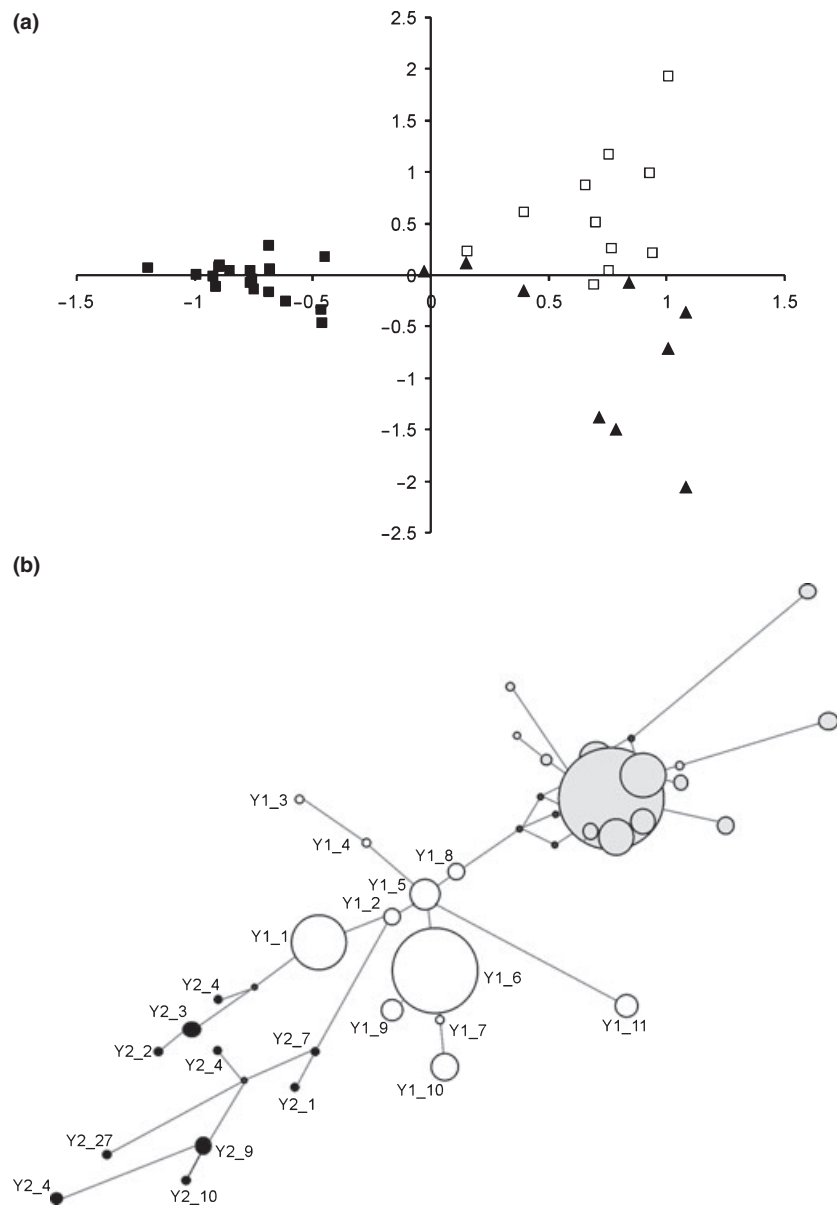


Figure 1 Graphical representations of genetic variability among Y1 and Y2 haplotypes. Plot a shows the dispersion of the observed 38 haplotypes calculated via correspondence analysis. Y1-haplotypes are in black triangles; Y2-haplotypes are in black squares except for those specific haplotypes found in African cattle, represented by open squares. Plot b shows a network tree constructed using the program NETWORK 4.5.2. The plot shows Y1-haplotypes (open circles), Y2-haplotypes (grey circles) except for those specific haplotypes for African cattle (black circles). Circles are proportional to haplotype frequencies. Regardless of the analysis carried out, the African-specific Y2-haplotypes tend to cluster nearer the Y1-haplotypes.

the model as the only hierarchical factor to be tested, it explained roughly 80% of the genetic variation. This variation was similar to the proportion of variation explained by 'Haplogroup' (71.09%) and 'Haplofamily' (74.51%). When 'Breed' was analysed within Continent or Geographical Area, it still accounted for more genetic variation than these latter two factors. The variation explained by the haplofamily was slightly (but significantly) higher than that explained by the haplogroup.

Discussion

Here we use a set of five cattle Y-specific microsatellites, including six loci, to assess diversity on the male path in taurine cattle. The microsatellite set used was fitted after

careful selection of markers (see Table S1). Our microsatellites basically coincide with those previously used by Ginja *et al.* (2009) in Portuguese cattle, and three of them (*INRA189*, *BM861* and *BYM-1*) were also used by Li *et al.* (2007) and Kantanen *et al.* (2009) in Ethiopian and Eurasian cattle respectively. The markers *INRA124*, *INRA 126* and *UMN0504* used in the aforementioned studies have been shown not to be male-specific here and in the previous analyses (Pérez-Pardal *et al.* 2009).

Haplotypic diversity found in our sample is similar to that of 0.064 (26 haplotypes among 405 bulls tested) reported by Kantanen *et al.* (2009), higher than that of 0.042 (13/307) reported by Ginja *et al.* (2009) in Portuguese cattle, but lower than that of 0.099 (16/161) found by Li *et al.* (2007) in Ethiopian cattle (including Holstein-Friesian).

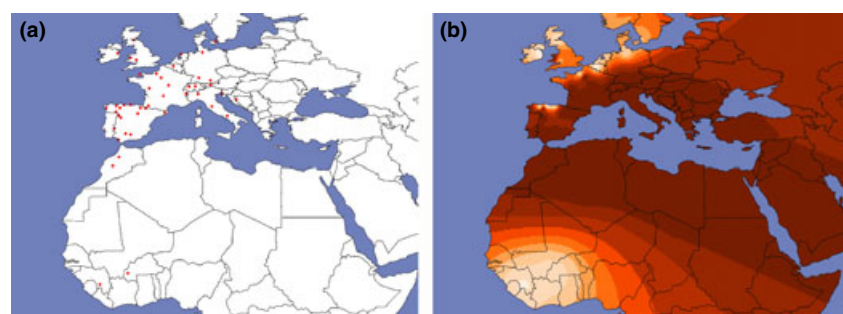


Figure 2 Locations of cattle populations sampled (Plot a) and synthetic maps illustrating geographical variation of the first principal component (Plot b) identified using principal component analyses (PCA). The projection of the main component of variation represents the main Y2 haplotypic subfamily, predominant in the cattle breeds sampled in Central Europe, the three European Peninsulas and Mediterranean Africa. By contrast, the Y1 haplotypes are the most frequent in the Atlantic European coastline, and a second Y2 haplotypic subfamily is typical of the Sub-Saharan cattle.

The results reported by Li *et al.* (2007) may be biased upwards as a result of the presence of taurine and zebu haplotypes in the Ethiopian samples, but also by the use of the putative fast-evolving marker *INRA126* that has been shown to have erratic behaviour with a high rate (55%) of parental incompatibility (Pérez-Pardal *et al.* 2009); it may be located on the pseudoautosomal region of the Y-chromosome (Li *et al.* 2007).

At the subspecies level, our study highlights the following: (i) haplogroups Y1 and Y2 previously identified by Götherström *et al.* (2005) have phylogenetic meaning; (ii) there is genetic variation within haplogroups that can be assessed using microsatellites; (iii) African cattle Y2-chromosomes can be differentiated from those of European cattle; (iv) there is clear geographical structuring in taurine cattle populations; and (v) this structuring is compatible with a poor overall genetic differentiation.

The fact that no haplotypes were shared between haplogroups clearly confirms the phylogenetic importance of the haplogroups identified by Götherström *et al.* (2005). They characterize two large cattle Y-chromosome lineages that are predominant in the European Atlantic coastline (Y1) and in the mainland European continent and Africa (Y2). However, in the case of cattle, biallelic markers, such as the Götherström *et al.* (2005) SNPs and indels, do not offer enough resolution to characterize Y-chromosome diversity. Y-specific microsatellites allowed at least two different Y2-lineages to be differentiated (Fig. 1). The African-specific

Y2-lineage clusters nearer to Y1 haplotypes than to European Y2-haplotypes. This suggests the existence of different wild sire populations with poor genetic differentiation. However, the set of markers used was not powerful enough to correctly discriminate among haplogroups. The identification of new Y-specific markers in cattle with higher resolution should confirm the genetic relationships between the different haplogroups.

Ginja *et al.* (2009) suggested a possible African introgression into Portuguese cattle based on one allele on marker *INRA189* previously found in N'Dama cattle (Edwards *et al.* 2000). Differences between West African genetic stock and European Y2-sires are more marked than previously shown. Here we demonstrate that these breeds have their own haplotypes that are not shared with any other Y2-cattle breeds. Haplotype diversity, adjusted for sample size, is quite similar for Y1 and Y2 haplogroups (0.055 ± 0.0006 and 0.067 ± 0.0002 , respectively). In any case, the higher haplotype diversity observed in Y2 cattle is mainly resulting from the variation observed in African Y2-haplotypes. The identified West African Y2-subfamily has an extremely high haplotype diversity 0.533 ± 0.1214 , whereas the other Y2 samples had haplotype diversity of 0.049 ± 0.0007 . Although our analyses do not allow us to reject the influence of other factors such as differences in population sizes or introgression from other populations, this fact suggests that West African cattle gather genetic variation of local origin. Summarizing the

Geographical area	1	2	3	4	5	6	7
1. Iberian Peninsula	0.629						
2. Atlantic Europe	0.363	0.775					
3. Continental Europe	0.736	0.353	0.913				
4. Italian Peninsula	0.704	0.328	0.873	0.892			
5. Balkan Peninsula	0.694	0.426	0.849	0.808	0.806		
6. Mediterranean Africa	0.664	0.322	0.814	0.780	0.756	0.753	
7. Sub-Saharan Africa	0.284	0.315	0.249	0.230	0.246	0.253	0.574

Table 3 Within-population (on diagonal) and between-population genetic identities (below diagonal) calculated by geographical area grouping of samples.

Table 4 Geographical structuring assessed using Analysis of Molecular Variance (AMOVA). 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. The statistical significance of variance components and Φ indices is evaluated by bootstrapping using 1000 replications. All estimates were statistically significant for $P < 0.0001$.

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}
Continents (Europe and Africa)	19.95	—	80.05	1.569	0.199	—	—
Geographical area ¹	46.95	—	53.05	1.486	0.469	—	—
Breed	79.76	—	20.24	1.328	0.798	—	—
Haplogroup (Y1 and Y2)	71.09	—	28.91	2.128	0.711	—	—
Haplofamily ²	74.51	—	25.49	2.106	0.745	—	—
Breeds within continents	11.55	70.34	18.11	1.484	0.819	0.795	0.115
Breeds within geographical areas ²	39.18	42.61	18.21	1.476	0.818	0.701	0.392

¹Geographical areas fitted were as follows (see Materials and methods and Table S1): Atlantic Europe (including the UK breeds, Dexter, Danish Red, Jersey, European and Holstein Friesian, Blanc-Bleu Belge and Normande cattle breeds); Iberian Peninsula (including Spanish breeds); Italian Peninsula (including Piemontese, Maremmana and Chianina breeds); Balkan Peninsula (including Croatian Simmental and Istrian cattle breeds); European mainland (including the rest of the European cattle breeds); Mediterranean Africa (including Bovine d'Oulmes and Bovine de Tidili samples); and Sub-Saharan Africa (including Lobi and N'Dama cattle samples).

²The haplofamilies fitted here are: (i) Y1-haplotypes; (ii) African-specific Y2 haplotypes; and (iii) the remaining Y2 haplotypes.

available genetic and archaeological evidence, Bradley & Magee (2006) suggested that African taurine cattle are not a simple subset of that domesticated in the Near East. This would be supported by the qualitatively different nature of their mtDNA composition, showing a large proportion of unique haplotypes (Troy *et al.* 2001), and that there is a significant component of microsatellite variation within African cattle that seems to be indigenous in origin (Hanotte *et al.* 2002). Though sample size from West African cattle is limited (only 17 samples), the evidence summarized by Bradley & Magee (2006) can also be applied to the Y-chromosome scenario identified here. In this respect, our findings would be consistent with a putative domestication event in Africa, including the use of local Y2-like wild sires. Moreover, the presence of specific African Y-chromosome haplotypes was found neither in Iberian cattle nor in the cattle of the other Southern European Peninsulas. Beja-Pereira *et al.* (2006) found African T1 mtDNA haplotypes in Southern and Northern Iberian cattle breeds. The lack of presence of African Y-specific haplotypes in Iberian cattle does not support recent introgression events. Genetic relationships between Iberian and African cattle populations would predate the traditional route associated with the Moorish invasions and occupation (Beja-Pereira *et al.* 2006).

As previously shown by Götherström *et al.* (2005), the Y1 and Y2 haplogroups have a clear geographical structure, with Y1 samples basically restricted to the Atlantic Europe and British Islands. The analysed breeds usually show Y-chromosomes belonging to a single haplogroup and are, in general, homogeneous at the Y-chromosome variation level. In most cases, the presence of Y1 and Y2 haplotypes

in a given breed can be explained by its recent history. Examples are the process of introgression of Fleckvieh, Brown Swiss and Friesian cattle into the Asturian cattle breeds during the first half of the 20th century (Goyache 1995) or the presence of Y2-haplotypes in Danish Red cattle, which is known to be introgressed by Brown Swiss cattle (Sørensen *et al.* 2005); the Red and White (Friesian) introgressed into Simmental cattle in Croatia; and the Y1 haplotypes found in N'Dama samples may have originated from Devon cattle influencing the sampled population during the 1930s.

The genetic heterogeneity of the wild sires used to create modern European cattle is a question of much debate. Götherström *et al.* (2005) suggested a large genetic influence of the primitive European aurochs in the formation of the present cattle breeds in Europe. However, Edwards *et al.* (2007b) and Bollongino *et al.* (2008), using ancient DNA, suggested that the model of a rapid introduction into Central Europe of cattle domesticated in the Near East without significant crossbreeding with local wild cattle remains unchallenged. Svensson & Götherström (2008) also suggested that the geographical consistency of the Y-specific haplogroups identified in European cattle could simply be the result of drift processes that occurred just before and during the formation of modern cattle breeds. The current analysis cannot support the hypothesis proposed by Götherström *et al.* (2005) suggesting that haplogroup Y2 would result from the domestication events occurring in the Near East during the Neolithic period, whereas Y1 would result from introgression of local male aurochs into domestic cattle populations all over the Atlantic European coastline: Both the haplotypic diversity and the

within-population expected heterozygosity (the complementary value of genetic identity) across different parts of Europe (Table S2) are low and substantially the same. This dataset suggests that both European Y1 and Y2 haplotypic families are associated with small initial numbers of sires. This would be consistent with a single domestication process. This contrasts with the high haplotypic variability and relatively high expected heterozygosity assessed in West African samples, in which there may have been the use of local sires (through genetically related Y2 sires domesticated in the Near East). The lack of samples from the Near East (the only definite centre of taurine cattle domestication) does not allow us to check this hypothesis.

Although there is room for improvement, this has been the first attempt, to our knowledge, to ascertain Y-specific genetic diversity and structure in African and European taurine cattle. Y-specific microsatellites have provided information about differentiation between Y-haplotypic lineages and relationships between cattle populations and geographical areas. Increasing the number of markers and the extent to which their properties are understood will influence the reliability of the inferences. Further analyses, including of samples from the Near East, as well as from different African regions, will help to clarify the role of Africa in the domestication of taurine cattle.

Acknowledgements

This work was partially supported by the MICIN project no. CGL2005-03761/BOS and the FCT grant POCI/CVT/56758/2004. LP-P is supported by grant MICINN BES-2006-13545. ABP and SC are supported by FCT grants SFRH/BPD/38096/2007 and SFRH/BPD/26802/2006 respectively. APdeL is supported by USDA-Hatch project MIN-16-019. Authors thank Eduardo Miura (Unión de Criadores de Toros de Lidia, Madrid), Pedro J. Azor (Universidad de Córdoba, Spain), Juan José Arranz (Universidad de León, Spain), Dimitri Pirottin (Université de Liège), Davide de Petris (Università degli Studi di Firenze, Italy), Jordi Jordana (Universitat Autònoma de Barcelona), Thomas Lecomte (JVL Company, Belgium-Congo) and Beate Scherf (FAO, Rome, Italy) for their support and help with sampling.

References

- Balaresque P., Sibert A., Heyer E. & Crouau-Roy B. (2006) Unbiased interpretation of haplotypes at duplicated microsatellites. *Annals of Human Genetics* **71**, 209–19.
- Bandelt H.-J., Forster P., Sykes B.C. & Richards M.B. (1995) Mitochondrial portraits of human populations. *Genetics* **141**, 743–53.
- Bandelt H.-J., Forster P. & Röhl A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37–48.
- Beja-Pereira A., Caramelli D., Lalueza-Fox C. *et al.* (2006) The origin of European cattle: evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8113–8.
- Bishop M.D., Kappes S.M., Keele J.W. *et al.* (1994) A genetic linkage map for cattle. *Genetics* **136**, 619–39.
- 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/journal.pone.0003418>.
- Bradley D.G. & Magee D.A. (2006) Genetics and origins of domestic cattle. In: *Documenting Domestication. New Genetic and Archaeological Paradigms* (Ed. by M.A. Zeder, D.G. Bradley, E. Emshwiller & B.D. Smith), pp. 317–28. University of California Press, Berkeley-Los Angeles-London.
- Bradley D.G., MacHugh D.E., Loftus R.T., Sow R.S., Hoste C.H. & Cunningham E.P. (1994) Zebu-aurine variation in Y chromosome DNA: a sensitive assay for genetic introgression in West African trypanotolerant cattle populations. *Animal Genetics* **25**, 7–12.
- Edwards C.J., Gaillard C., Bradley D.G. & MacHugh D.E. (2000) Y-specific microsatellite polymorphism in a range of bovid species. *Animal Genetics* **31**, 127–30.
- Edwards C.J., Baird J.F. & MacHugh D.E. (2007a) Taurine and zebu admixture in Near Eastern cattle: a comparison of mitochondrial, autosomal and Y-chromosomal data. *Animal Genetics* **38**, 520–4.
- Edwards C.J., Bollongino R., Scheu A. *et al.* (2007b) Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings. Biological Sciences* **274**, 377–1385.
- Excoffier L., Laval G. & Schneider S. (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50.
- Ginja C., Telo da Gama L. & Penedo M.C.T. (2009) Y Chromosome haplotype analysis in Portuguese cattle breeds using SNPs and STRs. *Journal of Heredity* **100**, 148–57.
- Giovambattista G., Ripoli M.V., De Luca J.C., Mirol P.M., Lirón J.P. & Dulout F.N. (2000) Male-mediated introgression of *Bos indicus* genes into Argentine and Bolivian Creole cattle breeds. *Animal Genetics* **31**, 302–5.
- Götherström A., Anderung C., Hellborg L., Elburg R., Smith C., Bradley D.G. & Ellegren H. (2005) Cattle domestication in the near East was followed by hybridization with aurochs bulls in Europe. *Proceedings of the Royal Society B* **272**, 2345–50.
- Goyache F. (1995) *Influencia de efectos sistemáticos sobre los caracteres de importancia económica en la raza Asturiana de los Valles*. PhD Thesis. Universidad Complutense de Madrid, Madrid, Spain. ISBN: 84-8466-997-1 (2002).
- Gutiérrez J.P., Royo L.J., Álvarez I. & Goyache F. (2005) MolKin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. *Journal of Heredity* **96**, 718–21.
- Hanotte O., Tawah C.L., Bradley D.G., Okomo M., Verjee Y., Ochieng J. & Rege J.E. (2000) Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds. *Molecular Ecology* **9**, 387–96.
- Hanotte O., Bradley D.G., Ochieng J.W., Verjee Y., Hill E.W. & Rege J.E. (2002) African pastoralism: genetic imprints of origins and migrations. *Science* **296**, 336–9.

- Kantanen J., Edwards C.J., Bradley D.G. *et al.* (2009) Maternal and paternal genealogy of Eurasian taurine cattle (*Bos taurus*). *Heredity* doi: 10.1038/hdy.2009.68.
- Kappes S.M., Keele J.W., Stone R.T., McGraw R.A., Sonstegard T.S., Smith T.P., Lopez-Corrales N.L. & Beattie C.W. (1997) A second-generation linkage map of the bovine genome. *Genome Research* 7, 235–49.
- de Knijff P. (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *American Journal of Human Genetics* 67, 1055–61.
- Li M.H., Zerabruk M., Vangen O., Olsaker I. & Kantanen J. (2007) Reduced genetic structure of north Ethiopian cattle revealed by Y-chromosome analysis. *Heredity* 98, 214–8.
- Liu W.-S., Beattie C.W. & Ponce de León F.A. (2003) Bovine Y chromosome microsatellite polymorphism. *Cytogenetic and Genome Research* 10, 53–8.
- Mathias N., Bayes M. & Tyler-Smith C. (1994) Highly informative compound haplotypes for the human Y chromosome. *Human Molecular Genetics* 3, 115–23.
- Meadows J.R.S., Hanotte O., Drögemüller C., Calvo J., Godfrey R., Coltman D., Maddox J.F., Marzanov N., Kantanen J. & Kijas J.W. (2006) Globally dispersed Y chromosomal haplotypes in wild and domestic sheep. *Animal Genetics* 37, 444–53.
- Mezzelani A., Zhang Y., Redaelli L., Castiglioni B., Leone P., Williams J.L., Toldo S.S., Wigger G., Fries R. & Ferretti L. (1995) Chromosomal localization and molecular characterization of 53 cosmid-derived bovine microsatellites. *Mammalian Genome* 6, 629–35.
- Nei M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY.
- Pérez-Pardal L., Royo L.J., Álvarez I., Ponce de León F.A., Fernández I., Casais R. & Goyache F. (2009). Female segregation patterns of the putative Y-chromosome specific microsatellite markers INRA124 and INRA126 do not support their use for cattle population studies. *Animal Genetics* 40, 560–4.
- Sørensen A.C., Sørensen M.K. & Berg P. (2005) Inbreeding in Danish dairy cattle breeds. *Journal of Dairy Science* 88, 1865–72.
- Svensson E. & Götherström A. (2008) Temporal fluctuations of Y-chromosomal variation in *Bos taurus*. *Biological Letters* 4, 752–4.
- Troy C.S., MacHugh D.E., Bailey J.F., Magee D.A., Loftus R.T., Cunningham P., Chamberlain A.T., Sykes B.C. & Bradley D.G. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410, 1088–91.
- Vaiman D., Imam-Ghali M., Moazami-Goudarzi K., Guerin G., Grohs C., Leveziel H. & Saidi-Mehtar N. (1994) Conservation of a syntenic group of microsatellite loci between cattle and sheep. *Mammalian Genome* 5, 310–4.
- Y-Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research* 12, 339–48.
- Zegura S.L., Karafet T.M., Zhivotovsky L.V. & Hammer M.F. (2004) High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of native american Y chromosomes into the Americas. *Molecular Biology and Evolution* 21, 164–75.
- Zerjal T., Wells R.S., Yuldasheva N., Ruzibakiev R. & Tyler-Smith C. (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *American Journal of Human Genetics* 71, 466–82.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Description of microsatellites tested in this study.

Table S2 Cattle Y-chromosome haplotypes defined by six Y-specific loci and their frequencies across breeds.

Table S3 Between breeds genetic identity computed from six Y-specific loci. Eigenvectors corresponding to each of the three principal components computed via Principal Component Analysis are also given for each breed.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery but or not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.