ORIGINAL ARTICLE

Reduced genetic structure of north Ethiopian cattle revealed by Y-chromosome analysis

MH Li¹, M Zerabruk², O Vangen², I Olsaker³ and J Kantanen¹¹Biotechnology and Food Research, MTT Agrifood Research Finland, Jokioinen, Finland; ²Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Norway and ³Department of Basic Sciences and Aquatic Medicine, Norwegian School of Veterinary Science, Oslo, Norway

Ethiopia is considered to be a putative migratory corridor for both Near-East *Bos taurine* and Arabian and Indian *B. indicus* cattle into East Africa. African pastoralism, which is associated with adaptation to specific habitats and farming systems, has contributed to the composite constitution of Ethiopian cattle. We analyse, for the first time, five Y-chromosome microsatellite markers from seven north Ethiopian cattle populations, using a European Holstein-Friesian population as a reference, to assess the paternal gene pool and to explore the mechanisms behind the genetic structure. Our results reveal that the indicine alleles predominate in the present populations, with only one animal in the Arado carrying the taurine alleles. The north Ethiopian cattle populations with one exception (Abergelle) are characterized by a general low Y-chromosome haplotype diversity, as well as by a reduced interpopulation variance ($\Phi_{ST} = 4.0\%$),

which can be a result of strong male-mediated selective sweeps. Population structure revealed by multidimensional-scaling analysis differentiates two populations (Arado and Abergelle) from the rest. Analysis of molecular variance does not lend support to the traditional classification for the populations, which is mainly based on physical characteristics. A network analysis indicates two closely related founding haplotypes accounting for a large proportion (50.0% in Abergelle and 85.0–94.7% in others) of north Ethiopian cattle Y-chromosomes. Our findings point to a common, but limited, paternal origin of the north Ethiopian cattle populations and strong male-mediated gene flow among them. The findings also provide insight into the historical immigration of cattle into East Africa.

Heredity (2007) **98**, 214–221. doi:10.1038/sj.hdy.6800931; published online 10 January 2007

Keywords: Y-chromosome; microsatellite; north Ethiopia; cattle

Introduction

Population genetics has emerged as a powerful tool for unraveling the history of cattle domestication. Unlike mitochondrial (Loftus *et al.*, 1994; Bradley *et al.*, 1996) and autosomal (MacHugh *et al.*, 1997) DNA markers, the Y-chromosome microsatellite is inherited patrilineally and provides evidence of paternal origins. Polymorphism at the Y-chromosome microsatellite can therefore reveal another perspective in the study of cattle domestication and breed development. Cattle Y-chromosome DNA polymorphism studies appeared in the early 1990s (Bradley *et al.*, 1994; Gwakisa *et al.*, 1994; Kemp and Teale, 1994). Surveys of variation in the Y-chromosome DNA to explore differentiation between taurine (*Bos taurus*) and zebu (*B. indicus*) breeds quickly followed (Hanotte, 1997; Edwards *et al.*, 2000). To date, only the *INRA124* microsatellite polymorphism has been examined for the Y-chromosome in indigenous African cattle breeds (Hanotte *et al.*, 2000). Of the five Y-chromosome microsatellites studied here, three loci (*INRA124*, *INRA189* and *BM861*) displayed the putative

taurine- and zebu-specific alleles, which can be useful indicators of male-mediated gene flow in hybrid populations (Edwards *et al.*, 2000).

The African continent is home to more than 150 different cattle breeds/populations classified into four broad categories of taurine, zebu, sanga and zenga (zebu-sanga) (Rege, 1999; Rege and Tawah, 1999). The modern African cattle breeds represent a unique genetic resource for improving livestock productivity on a global basis (Hanotte *et al.*, 2002). Despite the earliest cattle in Africa being undoubtedly *B. taurus* in nature (Bradley *et al.*, 1998), waves of immigration of humped *B. indicus* have profoundly changed the genetic landscape of African cattle (MacHugh *et al.*, 1997; Hanotte *et al.*, 2000, 2002). Zebu cattle are more recent immigrants in Africa ca. 2000–3000 BP (Clutton-Brock, 1989) with their major influence being centered in East Africa, around Ethiopia and neighbouring countries (Rege, 1999; Hanotte *et al.*, 2002). Thus, as one of the African regions with the strongest influence of intercontinental migrations and dense populations of cattle, north Ethiopian cattle populations may have a genetic structure shaped by several introductions of zebu, as well as by introgression of the taurine from the Near-East. It is generally agreed that the first zebras were brought to Ethiopia through Somalia by Semitic peoples from Arabia. Subsequent interbreeding of the zebras with the local taurine longhorns produced the present-day sanga cattle

Correspondence: Dr J Kantanen, Biotechnology and Food Research, MTT Agrifood Research Finland, H2-Building, FIN-31600 Jokioinen, Finland.
E-mail: juha.kantanen@mtt.fi

Received 9 May 2006; revised 26 September 2006; accepted 24 November 2006; published online 10 January 2007

(Rege, 1999). The second introduction of zebu led to the emergence of zenga breeds and their different strains adapted to the diverse ecological environments in the East African highlands. Even though there are no institutionalized schemes for genetic improvement of the breeds, exchange of superior bulls among closely related herdsmen is a common practice (Zerabruk and Vangen, 2005). The classification of breeds into zebu, sanga and zenga is by physical characteristics, geographical location and adaptive attributes (Rege, 1999).

The present study consists of one European Holstein-Friesian reference breed and seven cattle populations from north Ethiopia, representing the four broad categories described by Rege (1999) and by Zerabruk and Vangen (2005) (Table 1). We studied the variation of five Y-chromosome microsatellites, describing the distributions of haplotype diversity within each population. In addition, to better understand population relationships in a context of the traditional broad categories, we investigated patterns of affinity and diversification among the seven north Ethiopian cattle populations. The results may provide a genetic basis for the formulation of programs aimed at the conservation and sustainable use of unique cattle genetic resources in the region.

Materials and methods

Samples

DNA samples from 130 male cattle belonging to seven populations from north Ethiopia were analysed. Five of the seven populations were definitely allocated to three of Rege's (1999) four broad African cattle categories as follows: the humped zebu (Begaït), the zenga (Arado, Fogera) and the sanga (Afar, Raya). In addition, a recent on-farm investigation reported that the Abergelle and Irob cattle are classified as zenga type by the local farmers (Zerabruk and Vangen, 2005). Sample sizes and geographical distributions are given in Table 1 and Supplementary Figure S1. As a representative modern European taurine breed, the Holstein-Friesian was added to the analysed breeds as a reference. In all cases, particular efforts were made, using both pedigree information and the knowledge of local herdsmen, to ensure that animals were unrelated.

Y-chromosome microsatellite typing

Five Y-chromosome microsatellites (*INRA124*, *INRA126*, *INRA189*, *BM861* and *BYM-1*) were described previously by Edwards *et al.* (2000) and Ward *et al.* (2001). The structure of the loci are shown in Supplementary Figure S2. They were typed using 20–30 ng of DNA in PCR protocols on an MJ Research PTC 200 Peltier Thermal Cycler (MJ Research Inc., MA, USA) with a hot lid: *INRA124* (Vaiman *et al.*, 1994), *INRA126* (Vaiman *et al.*, 1994), *INRA189* (Kappes *et al.*, 1997) and *BM861* (Bishop *et al.*, 1994) were according to Edwards *et al.* (2000), and *BYM-1* (Matthews and Reed, 1991) according to Ward *et al.* (2001). The size characterization of PCR products was carried out on a Pharmacia ALF-express Automated Sequencer for loci *INRA124*, *INRA126*, *INRA189* and *BM861*, and on an Amersham MegaBACE 500 DNA Sequencer for locus *BYM-1*.

Table 1 Population information and allele frequencies of five Y-chromosome microsatellites in the seven North Ethiopian and one European Holstein-Friesian cattle populations

Group ^a	Breed/population	N ^b	INRA124					INRA189					BM861					INRA126					BYM-1				
			130 ^c (22 ^d)	132 (23)	88 (17)	90 (18)	98 (22)	156 (15)	158 (16)	182 (9)	184 (10)	186 (11)	188 (12)	254 (19)	256 (20)	258 (21)	260 (22)										
Taurine	Holstein-Friesian	31	—	1.00	—	—	1.00	—	1.00	0.0968	0.3548	0.5161	0.0323	—	0.0645	0.9355	—										
Zebu	Begaït	21	1.00	—	1.00	—	—	1.00	—	0.4286	—	0.5714	—	0.0476	0.9048	—	0.0476										
Sanga	Afar	18	1.00	—	0.9444	0.0556	—	1.00	—	0.7222	—	0.2778	—	—	0.9444	0.0556	—										
	Raya	14	1.00	—	0.9286	0.0714	—	1.00	—	0.50	—	0.50	—	—	1.00	—	—										
Zenga	Arado	20	0.95	0.05	0.85	0.10	0.05	0.95	0.05	0.60	—	0.40	—	—	0.95	0.05	—										
	Fogera	19	1.00	—	0.9474	0.0526	—	1.00	—	0.5789	—	0.4211	—	—	1.00	—	—										
	Abergelle	20	1.00	—	0.60	0.40	—	1.00	—	0.50	0.05	0.45	—	0.10	0.90	—	—										
	Irob	18	1.00	—	1.00	—	—	1.00	—	0.3333	—	0.6111	0.0556	—	1.00	—	—										

^aRege (1999) and Zerabruk and Vangen (2005).

^bNumber of samples.

^cBase pairs.

^dRepeat units.

Statistical analyses

Allele frequencies for all loci were determined by standard gene counting. In an initial pairwise analysis, allele frequencies at the five loci were compared between all pairs of populations using the Fisher exact test-based, genic-comparison option included in GENEPOP version 1.2 (Raymond and Rousset, 1995). Haplotype assignments were based on typing results. Haplotype diversities and their standard errors were calculated as Nei's (1987) unbiased estimates.

The software package Arlequin version 2.000 (Schneider *et al.*, 2000) was used to calculate several population genetic parameters, including pairwise genetic distances, analysis of molecular variance (AMOVA) and mismatch distribution. Significance levels of the variance components and Φ statistics were obtained by comparison of the actual values with the distribution of 10 000 values obtained by randomisation. Pairwise R_{ST} genetic distances between populations were computed as a linearization of Φ_{ST} values, that is, $D = \Phi_{ST}/(1 - \Phi_{ST})$ (Slatkin, 1995). Φ_{ST} , the F_{ST} analogue, counts the number of mutational steps separating each pair of haplotypes (Michalakis and Excoffier, 1996). The pairwise R_{ST} values were represented in two-dimensional space with multi-dimensional scaling (MDS) in the SPSS version 7.0 software package.

In AMOVA, two hierarchical levels (i.e., individuals clustered into populations and populations clustered into groups) were considered as the seven north Ethiopian cattle populations were subdivided into statistical groups according to criteria of either physical characteristics and historical information (Rege, 1999; Zerabruk and Vangen, 2005) or R_{ST} genetic distances. Φ statistics were then calculated, representing the haplotype correlations at the various levels of the hierarchical groupings: Φ_{CT} , cluster of the subpopulations relative to total populations; Φ_{SC} , subpopulation relative to cluster of the subpopulations; and Φ_{ST} , subpopulation relative to total populations. A median-joining network connecting different haplotypes was constructed using NETWORK 4.1 software (Bandelt *et al.*, 1999). All the 16 haplotypes found in this study are included in the analysis. According to Hurles *et al.* (2002) and Zegura *et al.* (2004), a reduced median analysis is performed first, and then the output is used as input for the median-joining network. Locus-specific weights are given proportional to their within-population variance components, so that loci with the highest variances are given the lowest weights. Hence, *INRA124*, *INRA126*, *INRA189*, *BM861* and *BYM-1* are given weights of 10, 9, 7, 10 and 9, respectively.

Results

Frequencies of the Y-chromosome microsatellite alleles in each population are given in Table 1. The indicine alleles predominate, comprising 99.2% (*INRA124/22*), 89.2% (*INRA189/17*) and 99.2% (*BM861/15*) of the observed alleles at the locus. Of the seven north Ethiopian cattle populations, three taurine alleles (*INRA124/23*, *INRA189/22* and *BM861/16*) were observed only in one Arado bull. *BYM-1* has a unimodal distribution with one frequent allele (*BYM-1/20*) and with the less frequent alleles differing from the most-frequent allele by a single repeat. *INRA126* is the locus with a bimodal allele-frequency

distribution. The genetic differences among north Ethiopian cattle are reflected by the possible 21 pairwise allele-frequency comparisons between the populations. Of the five Y-chromosome loci, *INRA189* shows highly significant differences ($P < 0.01$) in two comparisons and *INRA126* shows significant difference ($0.01 < P < 0.05$) in one comparison, whereas for *INRA124*, *BM861* and *BYM-1* no comparisons are statistically significant. In total, there are seven statistically significant comparisons at a level of $P < 0.05$ at all five loci, six of which arise from the differences between Abergelle and other breeds.

A total of 11 indicine and one taurine haplotypes, which are determined by the five Y-chromosome microsatellites, are observed in north Ethiopian bulls. Haplotype frequencies in each population are given in Table 2. Haplotype diversity varies from 0.527 (± 0.058) in Irob to 0.795 (± 0.036) in Abergelle. With the exception of Abergelle (0.795), the haplotype diversity values of north Ethiopian cattle populations (0.527–0.636) are lower than that of the European Holstein-Friesian (0.645) (Table 2). The most common haplotypes (H1 and H2) are widespread among north Ethiopian cattle populations. On the other hand, they represent a relatively low proportion in Abergelle cattle (50%), whereas their joint proportions are higher (85.0–94.7%) in all the other populations in the region. Haplotypes H3 and H4 are scattered such that haplotype H4 is shared by Arado, Abergelle and Fogera, whereas haplotype H3 is shared by Arado, Abergelle, Afar and Raya (Supplementary Figure S1).

Frequencies of Y-microsatellite haplotypes and the molecular differences were used to compute population genetic distances as Φ_{ST} (Supplementary Table S1). The highest interpopulation variance is observed between European Holstein-Friesian and African populations, as would be expected from their history. Among the north Ethiopian cattle, the Abergelle can be clearly differentiated from all the other populations. Pairwise values of Φ_{ST} show that, in some cases, neighbouring populations (e.g. Irob and Afar) are significantly different, whereas, in other cases, geographically distant populations (e.g. Irob and Fogera) have nonsignificant pairwise Φ_{ST} values. On the basis of the linearized Φ_{ST} distances, a picture on the relationships between populations was obtained by multidimensional scaling (MDS) analysis (Figure 1). A good fit between the two-dimensional plot and the source data (pairwise values of Φ_{ST}) was obtained, demonstrated by the low stress value (0.136) obtained (Zerjal *et al.*, 2002). As expected, the Holstein-Friesian population appears fairly separated from all the African cattle, evidencing the sharp dichotomy between the Y-chromosome pools of taurine and zebu. The shared haplotypes H3 and H4 in Arado and Abergelle contribute to their closeness, although the breeds are somewhat separated. The close grouping of the other five African cattle can be readily understood from the distribution and frequencies of haplotypes H1 and H2 (Table 1, Supplementary Figure S1).

The level of population structure was further assessed by estimating various Φ statistics by means of AMOVA (Excoffier *et al.*, 1992). The overall Φ_{ST} value calculated for the entire African sample, comprising the seven populations without grouping, is 0.040, indicating that quite a small proportion of the overall Y-chromosome variation results from interpopulation differences (Table 3). When the hierarchical approach is taken,

Table 2 Haplotype (INRA124-INRA189-BM861-INRA126-BYM-1) frequency and diversity ($\pm$ standard errors) in the seven North Ethiopian and one European Holstein-Friesian cattle populations

Haplotype (Repeat units)	Number of individuals							Total
	Begiti	Arado	Fogera	Afar	Raya	Abergelle	Irob	Holstein-Friesian
H1 (22-17-15-11-20)	11	7	7	5	7	3	11	—
H2 (22-17-15-9-20)	8	10	11	11	6	7	6	—
H3 (22-18-15-9-20)	—	1	—	1	1	2	—	—
H4 (22-18-15-11-20)	—	1	1	—	—	5	—	—
H5 (22-17-15-10-20)	—	—	—	—	—	1	—	—
H6 (22-18-15-9-19)	—	—	—	—	—	1	—	—
H7 (22-17-15-11-19)	—	—	—	—	—	1	—	—
H8 (22-17-15-9-21)	—	—	—	—	—	—	—	—
H9 (22-17-15-11-22)	1	—	—	1	—	—	—	—
H10 (22-17-15-9-19)	1	—	—	—	—	—	—	—
H11 (22-17-15-12-20)	—	—	—	—	—	—	—	—
H12 (23-22-16-9-21)	—	—	—	—	—	—	1	—
H13 (23-22-16-11-21)	—	1	—	—	—	—	—	3
H14 (23-22-16-10-21)	—	—	—	—	—	—	—	16
H15 (23-22-16-10-20)	—	—	—	—	—	—	—	9
H16 (23-22-16-12-21)	—	—	—	—	—	—	—	2
Diversity	0.590 (± 0.045)	0.636 (± 0.050)	0.541 (± 0.049)	0.559 (± 0.069)	0.582 (± 0.046)	0.795 (± 0.036)	0.527 (± 0.058)	0.645 (± 0.044)
								0.751 (± 0.015)

populations are divided into the three categorical groups described by Rege (1999) and Zerabruk and Vangen (2005) or three clusters based on the above MDS analysis. In the first case, the amount of variation among groups is not detected ($\Phi_{CT} = -0.019$, $P = 0.011$). However, the Φ_{ST} value for zenga indicates the existence of an internal population structure within the group (zenga, $\Phi_{ST} = 0.054$, $P < 0.001$). In the second grouping by MDS clustering, the Φ_{CT} is much higher than that obtained for traditional groups and is statistically significant ($\Phi_{CT} = 0.056$, $P < 0.001$). Thus, it would appear that, in north Ethiopian cattle populations, traditional classification mainly stemming from physical characteristics is not a reliable predictor of paternal lineages.

Mismatch distributions for Y-chromosome microsatellites are shown in Figure 2. The total north Ethiopian cattle conferred a multimodal model of mismatch distribution. The distribution had two substantial peaks, one at 0 and the other at 2, being likely in the context of a range expansion. The shape reflects the fact that in the north Ethiopian cattle populations most Y-chromosomes belong to two haplotypes, whereas other haplotypes occur at lower frequencies. The peak at 0 difference corresponds to the comparisons between individuals that share the same allele or have the same repeat units, and the second peak located at 2 differences represents two mutational steps separating the most frequent haplotypes. The additional peaks were at 8 and 10 differences, resulting from the presence of divergent taurine haplotype H12, which differs from the most frequent indicine haplotypes H1 and H2 by eight mutational steps.

Figure 3 displays a median-joining network for the 16 haplotypes detected in the seven north Ethiopian cattle populations and European Holstein-Friesian. This network exhibits two major distinct clusters (I, II) of closely related haplotypes, showing the clear grouping of taurine and indicine haplotypes. Cluster I consists exclusively of taurine haplotypes. The taurine haplotypes exhibit more than six mutational step differences from the indicine haplotypes and no haplotype, but a median vector, represents the connection between the two clusters, reflecting a marked divergence between them. The most striking feature of cluster II is the clear differentiation between those haplotypes with a motif INRA126/9 (cluster IIa) and the haplotypes with motifs INRA126/11 and INRA126/12 (cluster IIb). Cluster IIa and cluster IIb are connected by haplotype with INRA126/10. Cluster II is, therefore, characterized by the subclusters based on the repeats within locus INRA126. On the other hand, the similarity of the African indicine haplotypes is emphasized. Ten out of 12 indicine haplotypes are connected to the large central nodes (haplotypes H1 and H2) by a single mutation. The nodes may represent the ancestral indicine Y-chromosomes in the African cattle populations.

Discussion

This study aimed at understanding the genetic landscape of the Y-chromosome in north Ethiopian cattle populations and identifying the underlying domestication events that have led to the current distribution. The results also provide information about the genetic diversity within populations and the genetic affinities among populations from the paternal perspective.

Figure 1 Genetic relationships between populations based on multidimensional scaling (MDS) and a matrix of the pairwise R_{ST} genetic distances. Symbols used to identify the four broad categories are as follows: $\blacklozenge$, taurine; $\blacksquare$, zebu; $\blacktriangle$, sanga; $\diamond$, zenga.

Table 3 Analysis of molecular variance (AMOVA)

Sample	Number of groups	Number of populations	$\Phi_{ST}(P)$	$\Phi_{CT}(P)$	$\Phi_{SC}(P)$
North Ethiopian cattle	1	7	0.040 (<0.001)	—	—
<i>Traditional categorical groups</i>					
Zebu	1	1	—	—	—
Sanga	1	2	0.016 (0.042)	—	—
Zenga	1	4	0.054 (<0.001)	—	—
Overall	3	7	0.033 (<0.001)	-0.019 (0.011)	0.052 (<0.001)
<i>MDS groups</i>					
Abergelle	1	1	—	—	—
Arado	1	1	—	—	—
Irob, Raya, Begait Afar and Fogera	1	5	-0.023 (0.012)	—	—
Overall	3	7	0.053 (<0.001)	0.056 (<0.001)	-0.003 (<0.001)

Figure 2 Mismatch distributions for five Y-chromosome microsatellites in the total seven north Ethiopian cattle populations.

The indicine Y-chromosome predominates in the studied north Ethiopian breeds. Hanotte *et al.* (2000) and Rege and Bester (1998) concluded that East Africa is the cradle of the largest number of African zebu populations. The taurine alleles are very rare in north Ethiopian cattle and were only found in one Arado bull, which is most likely a result of recent crossbreeding or incomplete introgression of zebu patriline. The scarcity of the taurine Y-chromosome in north Ethiopia confirms the idea of its nearly total elimination due to male-mediated zebu influxes and expansions in the region (Bradley *et al.*, 1994, 1996; Hanotte *et al.*, 2000).

Y-chromosomal diversity was generally low in north Ethiopian cattle: with one exception (Abergelle), the haplotype diversity in north Ethiopian cattle is lower than that in European Holstein-Friesian. The reduced paternal haplotype diversity observed could be explained, on the one hand, by the introduction of Asian zebu with a limited number of bulls; on the other hand, the low haplotype diversities might also be due to the traditional breeding practice of sharing very few superior bulls. In addition, East African cattle, especially those distributed in semi-desert (Raya and Afar) and highland (Begait, Arado, and Irob) areas, have been regularly subjected to recurrent drought and epidemics, and may have gone through bottlenecks, which can cause losses also in paternal genetic variability. A reduced paternal genetic structure is supported by the facts that: (1) there is low Y-chromosome diversity, (2) only two closely related haplotypes encompass a large proportion (50.0% in Abergelle and 85.0–94.7% in others) of north Ethiopian cattle Y-chromosomes and (3) all other haplotypes are related to the two major haplotypes. Higher haplotype diversity is evident in Abergelle, where as many as three population-specific haplotypes are present. The Abergelle cattle are distinct in showing high frequencies of both haplotype H3, which is also found in Arado, Raya and Afar, and haplotype H4, which they share with

Figure 3 A reduced median-joining network of Y-chromosome microsatellite haplotypes in the seven north Ethiopian and one European Holstein-Friesian cattle populations.

Arado and Fogera. According to the recent field survey of Zerabruk and Vangen (2005), the Abergelle cattle are recognised by farmers to have strong adaptive advantages to the hotter and drier lowlands and the ability to cope with feed shortages and ticks during the long dry periods. All these features are favourably rated by the farmers. We speculate that milder bottleneck events affecting this population in lowlands might have contributed to the current situation.

The inferred pattern of interpopulation genetic affinities is illustrated by Φ statistics (Supplementary Table S1) and MDS analysis (Figure 1). As expected, the European Holstein-Friesian represents an obvious outlier with respect to north Ethiopian cattle, which is indicative of a sharp taurine-indicine dichotomy. As reflected in the haplotype frequency distribution (Supplementary Figure S1), a striking finding of MDS analysis is that Y-chromosomes of Abergelle and Arado are distinct from those of the other north Ethiopian cattle populations. An explanation for the clustering of Begait, Fogera, Raya, Afar and Irob would be their common paternal roots. The bulls of Abergelle cattle exhibit a number of characteristics, such as a fine skin, variable coat colour – black, black with white spots, chestnut, grey and light red, a small and not well-developed hump, and the especially small dewlap and naval flap as opposed to the typical features of the zebu breeds (Zerabruk and Vangen, 2005). These make Abergelle markedly distinct

from other East African cattle populations. Genetic analyses on autosomal markers also showed a differentiation of Abergelle from others (Zerabruk *et al.*, personal communication). The closeness between Abergelle and Arado might be supported by their geographical proximity.

Genetic differences among cattle populations can be correlated with traditional classification based on physical characteristics (Li *et al.*, 2005). However, the Φ_{CT} estimate for traditional groups is negative (-0.019 ± 0.011 ; Table 3). We would argue that traditional classification criteria alone cannot explain the substructure of Y-chromosome variation in north Ethiopian cattle populations. It has also been suggested that the topography, which contributes much to the wide variation in climate, soil, natural vegetation and settlement pattern of domestic species, has strongly influenced the distribution of cattle regionally and continentally (Kantaneu *et al.*, 2000; Hanotte *et al.*, 2002; Li *et al.*, 2002). The topography of the region is best described as a complex blend of highland (Arado), rugged terrain (Irob), lowland (Abergelle and Begait), steppe and semi-desert (Afra and Raya), and plain (Fogera). There seems to be no evidence for a correspondence between the inferred pattern of interpopulation genetic relationships and the topographically differentiated groups. However, as illustrated by Φ statistics (Supplementary Table S1), the population structure identified by the MDS analysis

seems to reveal the homogeneous groups better. It is interesting that the MDS reveals a special phylogeographic structure: the peripherally located populations fall into one group and the populations in between (Abergelle and Arado) have another genetic affinity, although they are also different from each other.

The network of haplotypes is shown in Figure 3. The effectiveness of the weighting is apparent in that the putative fast-evolving locus *INRA126* is represented by more than eight single-repeat links, whereas all the other loci are represented by one to two links. Three clusters have been identified: one far-separated taurine cluster and two indicine clusters that are linked to each other through a single haplotype. The Y-chromosome haplotype pattern supports a hypothesis that founding migrations of massive Asian zebu bulls have carried haplotypes H1 and H2 into East Africa. This is reflected in the structure of haplotypes (Figure 3) in that there are only two mutation steps of locus *INRA126* between the two founding indicine haplotypes, and the haplotypes and their derivatives are clustered by *INRA126* mutations. However, the scenario outlined here should be considered with caution. One potential caveat is that the fast-evolving locus *INRA126* might be located on the pseudoautosomal region of the Y-chromosome, which could explain its diversity. Moreover it would be desirable for more markers and additional breeds, including cattle from Arabian, Pakistani and Indian regions, to be studied to shed more light on the original source and migratory patterns of Asian zebu into East Africa.

Acknowledgements

We thank A Virta for the MegaBACE 500 runs and anonymous referees for helpful comments on the manuscript. The Norwegian State Educational Loan Fund (Lånekassen) and the NORAD/Mekelle University co-operation project are acknowledged for their financial contribution. The mobility scholarship funding granted by the Nordic Research Board to Merha Zerabruk to conduct part of this study at MTT Agrifood Research Finland is gratefully acknowledged. Dr Ilona Miceikiene from the K Janusauskas Laboratory of Animal Genetics of Lithuanian Veterinary Academy (Kaunas, Lithuania) organized part samples of the Holstein-Friesian population.

References

- Bandelt HJ, Forster P, Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Bio Evol* **16**: 37–48.
- Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SLF, Hawkins GA *et al.* (1994). A genetic linkage map for cattle. *Genetics* **136**: 619–639.
- Bradley DG, Loftus RT, Cunningham P, MacHugh DE (1998). Genetics and domestic cattle origins. *Evol Anthropol* **3**: 79–86.
- 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, MacHugh DE, Loftus RT, Sow RS, Hoste CH, Cunningham EP (1994). Zebu-aurine variation in Y chromosome DNA: a sensitive assay for genetic introgression in West African trypanotolerant cattle populations. *Anim Genet* **25**: 7–12.
- Clutton-Brock J (1989). Cattle in ancient north Africa. In: Clutton-Brock J (ed). *The Walking Larder: Pattern of Domestication, Pastoralism, and Predation*. Unwin Hyman: London. pp 200–214.
- Edwards CJ, Gaillard C, Bradley DG, MacHugh DE (2000). Y-specific microsatellite polymorphism in a range of bovid species. *Anim Genet* **31**: 127–130.
- Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.
- Gwakisa PS, Kemp SJ, Teale AJ (1994). Characterization of Zebu breeds in Tanzania using random amplified polymorphic DNA markers. *Anim Genet* **25**: 89–94.
- Hanotte O, Okomo M, Verjee Y, Rege E, Teale A (1997). A polymorphic Y chromosome microsatellite locus in cattle. *Anim Genet* **28**: 308–322.
- Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JEO (2002). African pastoralism: genetic imprints of origins and migrations. *Science* **296**: 336–339.
- Hanotte O, Tawah CL, Bradley DG, Okomo M, Verjee Y, Ochieng J *et al.* (2000). Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds. *Mol Ecol* **9**: 387–396.
- Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, Jobling MA (2002). Y chromosome evidence for the origins of oceanic-speaking peoples. *Genetics* **160**: 289–303.
- Kantanen J, Olsaker I, Holm LE, Lien S, Vilkkilä J, Brusgaard K *et al.* (2000). Genetic diversity and population structure of 20 north European cattle breeds. *J Hered* **91**: 446–457.
- Kappes SM, Keele JW, Stone RT, McGraw RA, Sonstegard TS, Smith TP *et al.* (1997). A second-generation linkage map of the bovine genome. *Genome Res* **7**: 235–249.
- Kemp SJ, Teale AJ (1994). Randomly primed PCR amplification of pooled DNA reveals polymorphism in a ruminant repetitive DNA sequence which differentiates *Bos indicus* and *B. taurus*. *Anim Genet* **25**: 83–88.
- Li MH, Sternbauer K, Haahr PT, Kantanen J (2005). Genetic components in contemporary Faroe Islands Cattle as revealed by microsatellite analysis. *J Anim Breed Genet* **122**: 309–317.
- Li MH, Zhao SH, Bian C, Wang HS, Wei H, Liu B *et al.* (2002). Genetic relationships among twelve Chinese indigenous goat populations based on microsatellite analysis. *Genet Sel Evol* **34**: 729–744.
- Loftus RT, MacHugh DE, Bradley DG, Sharp PM (1994). Evidence for two independent domestications of cattle. *Proc Natl Acad Sci USA* **91**: 2757–2761.
- MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG (1997). Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146**: 1071–1086.
- Matthews ME, Reed KC (1991). A DNA sequence that is present in both sexes of Artiodactyla is repeated on the Y chromosome of cattle, sheep and goats. *Cytogenet Cell Genet* **56**: 40–44.
- Michalakis Y, Excoffier L (1996). A genetic estimation of population subdivision using distances between alleles with special references for microsatellite loci. *Genetics* **142**: 1061–1064.
- Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.
- Raymond M, Rousset F (1995). GENEPOP (version 1.2): population genetic software for exact tests and ecumenicism. *J Hered* **86**: 248–249.
- Rege JEO (1999). The state of African cattle genetic resources I: classification framework and identification of threatened and extinct breeds. *Anim Genet Resour Inform* **25**: 1–25.

- Rege JEO, Bester J (1998). Livestock resource and sustainable development in Africa. In: Laurie P (ed). *Proceeding of 6th World Congress on Genetics Applied Livestock Production*. The University of New England: Armidale. pp 19–26.
- Rege JEO, Tawah CL (1999). The state of African cattle genetic resources II: geographical distribution, characteristics and use of present-day breeds and strains. *Anim Genet Resour Inform* **26**: 1–25.
- Schneider S, Kueffer JM, Roessli D, Excoffier L (2000). *Arlequin: A Software for Population Genetic Data Analysis*. Genetic and Biometry Laboratory, University of Geneva, Switzerland: Geneva.
- Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.
- Vaiman D, Mercier D, Moazami-Goudarzi K, Eggen A, Ciampolini R, Lepingale A *et al.* (1994). A set of 99 cattle microsatellites: characterization, syntenic mapping, and polymorphism. *Mamm Genome* **5**: 288–297.
- Ward TJ, Skow LC, Gallagher DS, Schnabel RD, Nall CA, Kolenda CE *et al.* (2001). Differential introgression of uniparentally inherited markers in bison populations with hybrid ancestries. *Anim Genet* **32**: 89–91.
- Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF (2004). High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of native American Y chromosomes into the Americas. *Mol Biol Evol* **21**: 164–175.
- Zerabruk M, Vangen O (2005). The Abergelle and Irob cattle breeds of North Ethiopia: description and on-farm characterisation. *Anim Genet Resour Inform* **36**: 7–20.
- Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002). A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am J Hum Genet* **71**: 466–482.

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