Original Article

Cytogenet Genome Res 120:150–156 (2008)
DOI: 10.1159/000118756

Cytogenetic and Genome Research

Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication

S. Hiendleder[a, b] H. Lewalski[c] A. Janke[d]

[a]Department of Molecular Animal Breeding and Biotechnology, Gene Center of the Ludwig Maximilian University, München (Germany); [b]JS Davies Animal Genetics and Epigenetics Group, Roseworthy Campus, School of Agriculture, Food & Wine, and Research Centre for Reproductive Health, The University of Adelaide, Adelaide (Australia); [c]Department of Reproductive Biology, University of Veterinary Medicine Hannover, Hannover (Germany); [d]Department of Cell and Organism Biology, Division of Evolutionary Molecular Systematics, University of Lund, Lund (Sweden)

Accepted in revised form for publication by M. Schmid, 21 December 2007.

Abstract. The taurine and zebuine cattle breeds comprise the majority of the world cattle population but their taxonomic status is still controversial. The two forms of cattle are currently classified as *Bos taurus* and *Bos indicus* species and are differentiated primarily by the presence or absence of a hump. However, these two species hybridize readily, producing fully fertile offspring. We have determined and analyzed complete *B. taurus* and *B. indicus* mitochondrial genome sequences to investigate the extent of sequence divergences and to study their taxonomic status by molecular dating. The sequences encompassed 16,338 and 16,339 nucleotides, respectively, and differed at 237 positions. Estimated divergence times indicated that the two cattle lineages separated 1.7–2.0 million years ago. Combined phylogenetic analyses of 18 new and 130 previously reported extant *B. taurus* and *B. indicus* control region sequences with data from 32 archaeological specimens of the extinct wild aurochs (*Bos primigenius*) identified four major maternal lineages. *B. primigenius* haplotypes were present in all but the *B. indicus* lineage, and one *B. taurus* sequence clustered with *B. primigenius* P haplotypes that were not previously linked with domestic cattle. The *B. indicus* cluster and a recently reported new *B. primigenius* haplotype that represents a new lineage were approximately equidistant from the *B. taurus* cluster. These data suggest domestications from several differentiated populations of *B. primigenius* and a subspecies status for taurine (*B. primigenius taurus*) and zebuine (*B. primigenius indicus*) cattle.

The tribe Bovini includes cattle (*Bos*) and buffalo (*Bubalus*) species that have been domesticated. Among the domesticated cattle species, wild ancestors still exist for yak (*Bos grunniens*), gayal/mithan (*Bos gaurus*) and Bali cattle (*Bos javanicus*), while taurine (*Bos taurus*) and zebuine cattle (*Bos indicus*) have only survived as domestic animals (Lenstra and Bradley, 1999). The latter two types of cattle account for the great majority of the world cattle population. The wild ancestor of *B. taurus* and *B. indicus*, the aurochs (*Bos primigenius*), ranged from the Pacific through Asia and Europe to the Atlantic, and from the northern tundra southwards into India and Africa. It is believed to have become extinct in Egypt as early as the 14th century B.C., but survived much longer in other regions, including Europe, where the last animal died in 1627 (Epstein and Mason, 1984).

Spontaneous or controlled interspecies hybridization among cattle appears to be common and has contributed to uncertainties about the origin and taxonomic status of domesticated forms. The male hybrids are usually sterile, but

This work was supported by a financial grant (HI 503/3-1, WO 685/3-1) from the Deutsche Forschungsgemeinschaft. S.H. is a JS Davies Professorial Fellow.

H.L. deceased in January 2005.

Request reprints from Professor Stefan Hiendleder
JS Davies Animal Genetics and Epigenetics Group, Roseworthy Campus
The University of Adelaide, Roseworthy SA 5371 (Australia)
telephone: +61 8 8303 7814; fax: +61 8 8303 7972
e-mail: Stefan.Hiendleder@adelaide.edu.au

1424–8581/08/1202–0150$24.50/0

females can breed successfully. The *B. taurus* and *B. indicus* cattle species present an exception in that they produce completely fertile male and female offspring (Gray, 1972). *Bos taurus* and *B. indicus* cattle differ in their Y chromosome structure (Potter and Upton, 1979; Goldammer et al., 1997) and studies of synaptonemal complexes in male *B. taurus* × *B. indicus* hybrids revealed elevated levels of meiotic chromosome pairing abnormalities (Switonski et al., 1990; Dollin et al., 1991a, b). However, the observed configurations and the levels of XY-autosomal associations and autosomal asynapsis were deemed unlikely to cause fertility problems (Switonski et al., 1990; Dollin et al., 1991b; Switonski and Stranzinger, 1998). This assumption is supported by reproductive parameters such as sperm motility, sperm concentration, and non-return rates of *B. taurus* × *B. indicus* hybrid males, which appear to be characterized by hybrid vigor effects (Thrift and Aaron, 1987).

Cattle breeds and strains have been morphologically differentiated primarily by the absence (*B. taurus*) or presence (*B. indicus*) of a hump (Lenstra and Bradley, 1999). More recently, nucleotide (nt) sequence variation in mitochondrial DNA (Loftus et al., 1994), microsatellites (MacHugh et al., 1997), satellite DNA (Nijman et al., 1999), and the Y chromosome (Kikkawa et al., 2003; Verkaar et al., 2003, 2004), of *Bos* species were employed to identify introgressions and to study population admixture. The molecular data have indicated a major taurine-zebuine dichotomy for cattle breeds in Europe and Asia, respectively, but suggested that African zebu cattle originated from earlier *B. taurus* strains by ancestral crossbreeding and male-mediated introgressions of *B. indicus* genetics (Bradley et al., 1996; MacHugh et al., 1997; Nijman et al., 1999). The mtDNA haplotypes recovered from archaeological specimens of European and Near Eastern aurochs and early domestic cattle have been interpreted as evidence for a Near Eastern origin of European *B. taurus* cattle that was followed by male-mediated European aurochs introgression (Troy et al., 2001; Götherström et al., 2005; Edwards et al., 2007).

The deep divergence between *B. taurus* and *B. indicus* mtDNA sequences points to two independent primary domestication events from genetically discrete aurochs strains, each possibly with a subspecies status (Loftus et al., 1994; Bradley et al., 1996). The latter studies assumed a minimum divergence time for the *Bos/Bison* split of 1 million years ago (Mya) based on fossil records but without giving specific details. From this, and evolutionary parameters taken from studies of the human mtDNA control region (CR), the evolutionary rate of an approximately 400 nt segment of the bovine CR was calculated for estimating cattle divergence times, including the *B. taurus* and *B. indicus* split at 0.2–1 Mya. Several studies on cattle domestication and diversity have based their divergence time estimates on this approach (MacHugh et al., 1997; Troy et al., 2001; Verkaar et al., 2004; Lai et al., 2006).

Complete mitochondrial genomes from artiodactyls, including the *B. taurus* and *B. indicus* sequences generated for this study, have now become available. These new data can be used to estimate the *B. taurus* and *B. indicus* divergence based on multiple, fossil based calibration points. Furthermore, dating algorithms that can account for rate differences along branches and include several calibration points of the molecular clock have been developed (Jobb et al., 2004). In the present study, we have used complete mitochondrial genome sequences to estimate the *B. taurus* and *B. indicus* divergence time, and analyzed mtDNA CR sequences to reconstruct and clarify the phylogeny of extant *B. taurus* and *B. indicus* in a *B. primigenius* context. Furthermore, the data also allow detailing the intraspecies mitochondrial genome variation with potentially functional relevance that is encountered in *B. taurus* and *B. indicus*, as well as their composite breeds.

Materials and methods

Sequencing of B. taurus and B. indicus mitochondrial DNA

We selected Brangus, Brown Swiss, Holstein, Simmental, and so called 'Dwarf' Zebu cattle for sequencing of the complete mtDNA control region (CR). The latter breed was founded with Zebu cattle from Sri Lanka. Total cellular DNA was isolated from blood leukocytes by standard proteinase K/phenol-chloroform procedures (Sambrook et al., 1989). The CR was amplified with primers mtDA (5′-CTCACCATCAACCCCCAAAGCT-3′) and mtDB (5′-TCATCTAGGCATTTTCAGTG-3′) using Taq-polymerase (Hybaid, Germany) as described (Hiendleder et al., 1999). We used PCR-RFLP analyses prior to sequencing in order to avoid sequencing identical haplotypes and digested amplicons with *Nla*III (New England Biolabs, Germany) (Brüggerhoff et al., 2002). The *B. indicus* haplotypes were identified according to *Bst*NI and *Dde*I restriction fragment profiles (Hiendleder et al., 2003). For sequencing, CRs were PCR amplified with Pwo-polymerase and cloned into pCR-TOPO-BluntII (Invitrogen, The Netherlands).

We selected a *B. taurus* mitochondrial genome from an animal of the Simmental breed and a *B. indicus* mitochondrial genome from a 'Dwarf' Zebu for complete nt sequence analysis. Mitochondrial DNA was purified from liver tissue (Hiendleder, 1993) and cloned into the plasmid pZErOTM-2 (Invitrogen). A 2.6 kb *Eco*RI/*Hin*dIII *B. taurus* mtDNA fragment was refractory to cloning and was amplified by PCR from purified mtDNA with a specific primer pair. Both strands of all cloned fragments and the single PCR amplified fragment were sequenced by a primer walking approach using standard procedures on a LICOR 4200 (MWG Biotech, Germany). Sequence coverage was 2–6 fold. Primer sequences are available on request.

Molecular dating and phylogenetic analyses

The protein coding gene alignment included our new complete mtDNA sequences of *Bos taurus* (AF492351) and *Bos indicus* (AF492350) cattle, the published *Bos taurus* reference sequence (Anderson et al., 1982; V00654), one additional complete *Bos indicus* sequence (NC_005971) and two additional complete *Bos taurus* sequences (DQ124389, DQ124403) from the database. The latter two *Bos taurus* sequences appeared much more diverged from other *Bos taurus* sequences in an initial explorative CR analysis (data not shown). To this set of cattle sequences, we aligned the protein coding mtDNA sequences from *Ammotragus lervia* (NC_009510), *Balaenoptera musculus* (X72204), *Bos grunniens* (NC_006380), *Bubalus bubalis* (NC_006295), *Capra hircus* (AF533441), *Cervus elaphus* (NC_007704), *Cervus nippon taiouanus* (NC_008462), *Elaphodus cephalophus* (NC_008749), *Hippopotamus amphibious* (NC_000889), *Muntiacus crinifrons* (AY239042), *Muntiacus muntjak* (AY225986), *Muntiacus reevesi* (AF527537), *Ovis aries* (AF010406), *Physeter macrocephalus* (AJ277029), *Rangifer tarandus* (NC_007703), and wild and domestic *Sus scrofa* (AJ002189, AF304202, AF304200, AF304201, AF304203). In order to root the tree in the phylogenetic analysis, the mtDNA sequence data of two Perissodactyla, *Equus caballus*, horse (X79547) and *Equus asinus*, donkey (X97337), were added to the alignment.

Table 1. Absolute nucleotide differences between complete *Bos taurus* and *Bos indicus* mitochondrial genomes

	Control region[b]			Intergenic	rDNA			tRNA	Protein coding[c]			
	subs	indels	all	subs	subs	indels	all	subs	1st	2nd	3rd	all
B. taurus A/*B. taurus* H[a]	4	0	4	0	0	2	2	0	1	2	2	5
B. taurus H/*B. indicus* H[a]	46	3	49	2	19	0	19	12	27	5	123	155

[a] *B. taurus* A, according to Anderson et al. (1982); *B. taurus* H, according to Hiendleder et al. (this study), *B. indicus* H, according to Hiendleder et al. (this study).
[b] subs: nucleotide substitutions; indels: insertions/deletions; all: combined nucleotide differences.
[c] 1st, 2nd, 3rd: first, second and third codon position, respectively.

Phylogenetic analysis was based on the alignment of complete CR sequences from six complete *B. taurus* and *B. indicus* cattle mitochondrial genomes listed above, and our 15 new *B. taurus* and *B. indicus* complete CR sequences. In addition, we included the following 130 complete or nearly complete *B. taurus* and *B. indicus* CR sequences from the database to the alignment: AF022916–AF022924, AF034438–AF034446, AF069432–AF069438, AF204778–AF204782, AF209120–AF209126, AF361452–AF361461, AY126697, BTU87633–BTU87650, BTU87893–BTU87905, BTU92230–BTU92244, L27712–L27737. To this alignment, 32 of the longest, approximately 400 nt long, fossil DNA sequences of *Bos primigenius* (Edwards et al., 2007) were added.

The protein-coding sequences were aligned manually and then concatenated. A χ^2 test as implemented in the TREE-PUZZLE program package (Strimmer and von Haeseler, 1996) was used to statistically evaluate differences in nucleotide (nt) and amino acid (aa) composition of the data set. Maximum likelihood (ML) analyses were done in TREEFINDER v May 2007 (TF) (Jobb et al., 2004) and TREE-PUZZLE v 5.2 (TP) (Schmidt et al., 2002). All nt analyses assumed the General Time Reversible (GTR) model of sequence evolution (Lanave et al., 1984) and rate heterogeneity with four classes of Γ distributed rate categories (Yang, 1994) and one class of invariable sites (4 Γ + I), as suggested by Modeltest version 3.7 (Posada and Crandall, 1998) for protein coding genes. The aa sequence analyses assumed the mtREV-24 (Adachi and Hasegawa, 1996) model of sequence evolution and 4 Γ + I. The observed *P* values and absolute number of substitutions were calculated by PAUP* (Swofford, 2002). The CR sequences were analysed by neighbour joining (NJ) and the TN-93 model of sequence evolution.

Divergence times were estimated with TF on nt ML trees provided by TF and based on local rate minimum deformation method (LRMD) (Jobb et al., 2004). Errors were calculated from ML branch lengths that were generated from 100 bootstrap sequence replicates and a fixed ML nt tree topology. The mean values and their standard deviations (s.d.) were recorded. The average *B. taurus*/*B. indicus* divergence time was used to derive mean intra-species divergence times based on their mean distance values.

Four reference points were used for calibrating the molecular clock. For three reference points a range of the possible oldest and possible youngest splits were chosen as the lower and upper limit for the calibration. For estimating divergence time with the LRMD method (Jobb et al., 2004) one divergence time needed to be fixed to one date. The cow and sheep divergence was set to 28 to 18 Mya (Hiendleder et al., 1998; Benton and Donoghue, 2007), the sperm whale and blue whale divergence was set to 35 to 30 Mya, and the whale and hippopotamus split was set to 54 to 52 Mya (Arnason and Gullberg, 1996; Bajpai and Gingerich, 1998; Arnason et al., 2000; van Tuinen and Hadly, 2004). For the fixed divergence time the split between whales and ruminants at 60 Mya was chosen, because this is the most narrowly defined one (Arnason et al., 1996).

Results

Sequence variation between complete B. taurus and B. indicus mitochondrial genomes

The length of the complete new *B. taurus* H and *B. indicus* H mitochondrial genome sequences is 16,338 and 16,339 nt, respectively. The *B. taurus* H sequence differs only at eleven nt positions from the currently used *B. taurus* A reference sequence established by Anderson et al. (1982), but differs at 237 nt positions from our *B. indicus* H reference sequence. The differences for specific regions of the mitochondrial genome between complete *B. taurus* and *B. indicus* sequences are detailed in Table 1.

Surprisingly, *B. taurus* A and *B. taurus* H differed with similar frequency at all three codon positions of the protein-coding sequences. This resulted in a high proportion of non-synonymous nt substitutions. Although the statistical basis is limited, more substitutions at the fast evolving third codon position than at the second and first codon positions were expected. This could indicate that the *B. taurus* reference sequence contains some, albeit few, sequencing artefacts. The predicted aa differences are Ala↔Gly (*COX3*), Thr↔Lys (*ND5*), and Val↔Ile (*CYB*) and it is noteworthy that all three aa are identical in *B. taurus* H and *B. indicus* H sequences. The two indels (insertions/deletions) in rDNA genes observed in the *B. taurus* A and H comparison are located in poly A and poly C tracts that might have been prone to sequence ambiguities. Again, the *B. taurus* H and *B. indicus* H sequences are identical in the respective poly A and poly C tracts.

The comparison of the CR sequences of the *B. taurus* H and *B. indicus* H sequences did not show any transversions. Sequence variation affected several CR elements described in Hiendleder et al. (2003). This includes the CSB2+3, LSP and O_H regions, TAS-A, and term, a termination associated sequence motif. One of the two substitutions in intergenic regions is between tRNAs asparagine and cysteine in the stem-loop structure of O_L. The 12S and 16S rDNA showed five and 14 substitutions, respectively, but no indels. Seven tRNAs (Leu, Asn, Cys, Asp, Gly, Glu, Pro) showed nt variations, with four substitutions detected in $tRNA^{Asp}$ alone. The ratio of substitutions at codon positions of protein coding genes was as expected. Approximately 10% of variable

nt positions resulted in predicted aa variations between *B. taurus* and *B. indicus* cattle. The observed differences affected seven respiratory chain enzyme subunits: CYB (Val↔Ile), ND1 (Thr↔Ile), ND3 (Thr↔Ala), ND4 (Asn↔Ser, Leu↔Val), ND5 (Phe↔Leu, Ile↔Thr, Tyr↔His, Ile↔Val, Tyr↔Phe, Asn↔Ser, Ile↔Met), ATP6 (Thr↔Ala, Thr↔Ala, Met↔Val), and ATP8 (Val↔Ile, Thr↔Ala).

Molecular dating of B. taurus and B. indicus divergence time

The alignment was straightforward and no gaps needed to be included inside the protein coding genes, except for a single, unambiguous aa deletion in ATP8 in the goat. However, some genes terminated earlier thus shortening the ATP8 in whales by 9 aa, and ND5 by 12 aa and 5 aa in the hippopotamus and in the sika deer, respectively. After removing the stop codons and trimming for length differences at the 3′ end, 11,373 nt remained for the phylogenetic analysis. The nt composition was non-homogenous in some species as tested by a χ^2 test analysis, while the aa composition was homogenous for all species. The nt ML tree and the branch lengths are shown in Fig. 1. The aa analysis indicated a different position of *Rangifer tarandus,* the reindeer: (*Elaphodus*, (*Rangifer*, *Cervus*)). However, the support from aa sequence analysis for either position of the genus *Rangifer* was low. The mean estimates of the divergence times based on the four calibration points and the LRMD method are shown in Table S1 (supplementary information, www.karger.com/doi/10.1159/000118756). The divergence of the *B. taurus* and *B. indicus* clusters has been estimated at 2.0 ± 0.14 Mya based on nt sequences and at 1.7 ± 0.50 Mya based on aa sequences. The split *B. taurus*/*B. grunniens* has been dated to 9.40 ± 0.50 Mya using nt and 5.16 ± 1.11 Mya using aa sequences (Table S1). The other divergence times generally agree with previously published data. Surprisingly, one of the additional two *B. taurus* sequences retrieved from the database, *B. taurus* H1, grouped with *B. indicus* sequences (Fig. 1).

Phylogenetic relationships between B. taurus, B. indicus and B. primigenius

The alignment of the complete or nearly complete CRs was 919 nt long. The combined phylogenetic analyses that included partial *B. primigenius* CRs identified four major maternal lineages: *B. indicus*, *B. taurus*, *B. primigenius* P, and *B. primigenius* E. The NJ tree shows the position of individual *B. taurus*, *B. indicus*, and *B. primigenius* sequences within each cluster (Fig. 2). The *B. taurus* cluster comprised haplotypes from taurine and African zebuine cattle and *B. primigenius* haplotypes recovered from archaeological specimens collected in Italy. The recently reported *B. primigenius* haplotype E represented a separate lineage that was approximately at the same distance from the *B. taurus* cluster as the *B. indicus* cluster. The CR data placed the divergent *B. taurus* H1 sequence (see above) closer to the main *B. taurus* cluster, but still distinctly outside that clade. The *B. taurus* FC3 sequence clustered with the *Bos primigenius* P haplotypes that were not previously linked with domestic

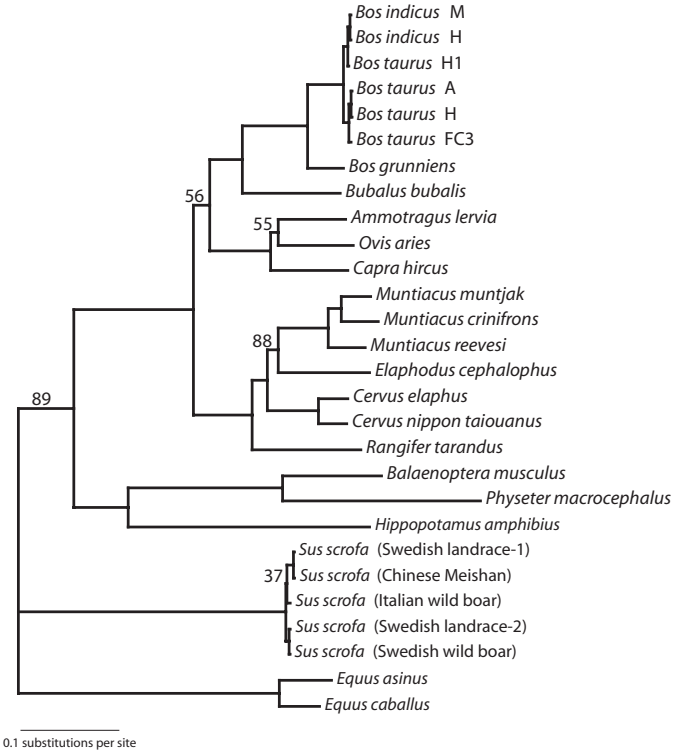

Fig. 1. Mitogenomic nucleotide sequence maximum likelihood (ML) tree (GTR + 4Γ + I) of artiodactyl and cetacean relationships. The ML bootstrap support is shown only for branches with <95% support.

cattle. We therefore excluded the *B. taurus* H1 and FC3 sequences from the divergence time estimation of the diversification of the *B. taurus* and *B. indicus* cluster (see below).

Divergence within B. taurus and B. indicus clusters

The resolution within the clusters was generally low because of the limited number of substitutions in each population. The average distance among *B. taurus* haplotypes was 0.00592 ± 0.0021 substitutions/site. The average distance between *B. taurus* and *B. indicus* was 0.052 ± 0.002 substitutions/site and their divergence time was estimated to be 1.997 Mya. Thus, the average divergence time for sequences in the *B. taurus* cluster is 227 thousand years ago (kya) (1.997/0.052 × 0.00592). The standard deviation (s.d.) for the *B. taurus* CR distances is 0.0021 and was taken to estimate the s.d. of the average *B. taurus* divergence to ±77 kya. The average distance between the *B. indicus* cluster sequences was 0.00525 ± 0.0023 substitutions/site. Thus, the divergence time among the *B. indicus* lineages was estimated at 202 ± 88 kya. The corresponding dates based on the aa divergence times of *B. taurus* and *B. indicus* (1.695 Mya) are about 20% younger. These dates refer to the most recent common ancestor of the respective lineages and are not associated with domestication events. The divergence times of *B. primigenius* P sequences were not cal-

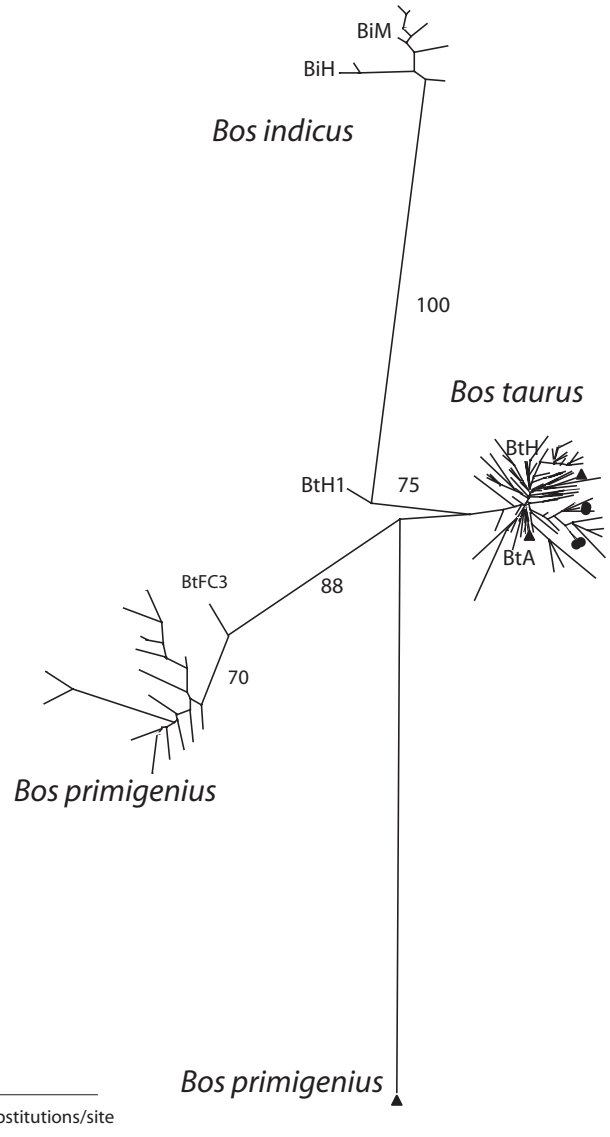

Fig. 2. Phylogenetic relationships among *B. taurus, B. indicus* and *B. primigenius* cattle based on neighbour joining analysis of control region sequences. Complete mitochondrial genomes are indicated. BtA: current *B. taurus* reference sequence (Anderson et al., 1982); BtH and BiH: new *B. taurus* and *B. indicus* reference sequences (Hiendleder et al., this study); BtH1, BtFC3, BiM: additional *B. taurus* (H1, FC3) and *B. indicus* (M) sequences from the database. Triangles and circles indicate sequences from *B. primigenius* and African zebuine cattle, respectively that were assigned to the *B. taurus* cluster. The *B. primigenius* haplotype E sequence, which represents a new lineage, is also marked with a triangle. Bootstrap values are shown for branches with >50% support.

culated in this way because of the short fossil DNA sequences and the corresponding poor statistics. However, the branch lengths (Fig. 2) indicate similar values for the *B. primigenius* cluster as for the *B. taurus* and *B. indicus* clusters.

Discussion

Reference sequences and variation in B. taurus and B. indicus mitochondrial genomes

Our new *B. taurus* mitochondrial genome sequence (AF492351) was highly similar to the currently used bovine reference sequence that was established more than 25 years ago (Anderson et al., 1982). We found only three amino acid substitutions, with some evidence for possible sequencing artefacts in the reference sequence, but the statistical basis for this assumption was limited (Table 1). The frequency of sequencing errors in the original reference sequences for the human (Anderson et al., 1981) and mouse mitochondrial genome (Bibb et al., 1981) was low and confirmed the high level of accuracy achieved in the early sequencing projects (Andrews et al., 1999; Bayona-Bafaluy et al., 2003).

Phylogenetic analyses assigned the bovine reference sequence, *B. taurus* A, and our new *B. taurus* H sequence, to *B. taurus* haplogroups (Troy et al., 2001) (Fig. 2). Overall sequence variation in the *B. taurus* cluster is apparently low, and results from our reference sequence comparison are in line with a PCR based study on mitochondrial genome variation in Japanese *B. taurus* cattle (Mannen et al., 2003). Additional comparisons between the *B. taurus* H sequence and the database sequence *B. taurus* FC3 (DQ124389), which grouped with the *B. primigenius* P haplotypes (Fig. 2), yielded a higher level of divergence and identified four variable aa. However, this was still considerably lower than the 17 predicted aa substitutions in the *B. taurus* H versus *B. taurus* H1 comparison. The latter database sequence (DQ124403, Korean Holstein-Friesian cattle) is enigmatic, as it was clearly assigned to the *B. indicus* cluster based on its aa sequence, but grouped near the *B. taurus* cluster by CR sequence analysis. The gene pool of some Asian cattle breeds is influenced by contributions from *B. grunniens* (Kikkawa et al., 2003) and *B. javanicus* (Kikkawa et al., 2003; Nijman et al., 2003). We excluded such an origin of the *B. taurus* H1 haplotype by our dating analyses (Fig. 1) and CR comparison (Nijman et al., 2003). More detailed analyses and comparisons of aa and nt sequence data (not shown) confirmed the peculiar nature of the H1 mitochondrial genome that presently cannot be explained.

The *B. indicus* H mitochondrial genome provides the first complete reference sequence for *B. indicus* mtDNA. Comparisons with *B. taurus* H revealed a large number of coding region sequence polymorphisms and predicted 17 aa substitutions, some of them non-conservative (Table 1). The number of non-synonymous substitutions is similar to those observed in intraspecies comparisons of mtDNAs that represent remote points in genealogical trees of mice and humans (Hiendleder, 1998). However, it is apparently higher than in sheep (Hiendleder, 1998) and lower than in pig (Kijas and Andersson, 2001). A comparison of our *B. indicus* H reference sequence with the *B. indicus* M sequence from the database (NC_005971) showed only a very limited amount of sequence variation that was in the order of *B. taurus* haplogroup differences. Because the *B. indicus* H reference sequence and the *B. indicus* M sequence belong to the

two identified *B. indicus* haplogroups I1 and I2 (Lai et al., 2006), respectively, it is possible that these mtDNAs represent the maximum in sequence variation to be encountered in *B. indicus* cattle.

Structural variation in mtDNA has been associated with specific phenotypes in various species, including human (http://www.mitomap.org) and mouse (Roubertoux et al., 2003). In cattle, mtDNA variants reportedly affect milk fat yield, carcass composition, and fertility traits (reviewed in Hiendleder, 1998; Hiendleder et al., 2005). The differences in *B. taurus* and *B. indicus* mitochondrial genomes could therefore contribute to the differences in phenotype between taurine and zebuine cattle.

Molecular dating and phylogenetic relationships between B. taurus, B. indicus and B. primigenius

Molecular dating based on solid fossil calibration points and algorithms estimated the *B. taurus*/*B. indicus* divergence at 1.7 ± 0.50 Mya based on aa sequences and at 2.0 ± 0.14 Mya based on nt sequences. These dates are clearly older than the previous estimate of 0.2–1.0 Mya based on an assumed *Bos/Bison* split at 1 Mya and evolutionary parameters taken from human mtDNA CR studies (Loftus et al., 1994; Bradley et al., 1996). The new divergence time estimates assign a clear subspecies status to the aurochs populations that appear to have been in an advanced state of speciation (Avise et al., 1998) before giving rise to domesticated forms of cattle. We therefore propose that the two domestic forms are named *B. primigenius taurus* and *B. primigenius indicus*.

The mtDNA haplotypes of European aurochs (P), and Near Eastern aurochs and early domestic cattle (T), have been interpreted as evidence for a Near Eastern origin of European *B. taurus* cattle (Troy et al., 2001; Edwards et al., 2007). This assumption was to a large extent based on the absence of *B. taurus* T haplotypes in European aurochs. The T and T1–T4 clusters were identified as central *B. taurus* haplogroups. T1 is frequent in Africa, T3 predominates in Europe and along with T and T2 comprises most of the Near Eastern variation (Troy et al., 2001), while T4 was found in East Asia (Mannen et al., 2004; Lai et al., 2006). However, T haplotypes reported from *B. primigenius* specimens in Italy (Fig. 2) suggested at least some maternal contributions from European aurochs to *B. taurus* breeds, and may indicate a more diverse and complex domestication history for taurine cattle (Beja-Pereira et al., 2006). The *B. taurus* FC3 sequence clustered with *B. primigenius* P haplotypes that were not previously linked with domestic cattle. This supports maternal contributions from P haplotype aurochs or a closely related population to *B. taurus* cattle and identifies three major mtDNA lineages among extant taurine and zebuine cattle.

A more complex genetic basis and domestication history that extends beyond the originally observed East-West dichotomy in mtDNA of most farm animal species (MacHugh and Bradley, 2001) is now increasingly evident. A prime example is the sheep, where the original two mtDNA lineages reported in various breeds (Hiendleder et al., 2002) have recently increased to five (Meadows et al., 2007). The position of the Neolithic *B. primigenius* haplotype E, and perhaps *B. taurus* H1, if confirmed, suggests a very diverse aurochs population that could have contributed to the gene pool of taurine and zebuine cattle.

GenBank Submissions

Complete *B. taurus* and *B. indicus* mtDNA control regions (AF492426–AF492440) and complete *B. taurus* and *B. indicus* mitochondrial genomes (AF492351 and AF492350) were submitted to GenBank.

References

Adachi J, Hasegawa M: Model of amino acid substitution in proteins encoded by mitochondrial DNA. J Mol Evol 42:459–468 (1996).

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, et al: Sequence and organization of the human mitochondrial genome. Nature 290:457–465 (1981).

Anderson S, de Bruijn MH, Coulson AR, Eperon IC, Sanger F, et al: Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J Mol Biol 156:683–717 (1982).

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N: Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23: 147 (1999).

Arnason U, Gullberg A: Cytochrome b nucleotide sequences and the identification of five primary lineages of extant cetaceans. Mol Biol Evol 13: 407–417 (1996).

Arnason U, Gullberg A, Janke A, Xu X: Pattern and timing of evolutionary divergences among hominoids based on analyses of complete mtDNAs. J Mol Evol 43:650–661 (1996).

Arnason U, Gullberg A, Gretarsdottir S, Ursing B, Janke A: The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates. J Mol Evol 50:569–578 (2000).

Avise JC, Walker D, Johns GC: Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc Biol Sci 265:1707–1712 (1998).

Bajpai S, Gingerich PD: A new Eocene archaeocete (Mammalia, Cetacea) from India and the time of origin of whales. Proc Natl Acad Sci USA 95: 15464–15468 (1998).

Bayona-Bafaluy MP, Acín-Pérez R, Mullikin JC, Park JS, Moreno-Loshuertos R, et al: Revisiting the mouse mitochondrial DNA sequence. Nucleic Acids Res 31:5349–5355 (2003).

Beja-Pereira A, Caramelli D, Lalueza-Fox C, Vernesi C, Ferrand N, et al: The origin of European cattle: evidence from modern and ancient DNA. Proc Natl Acad Sci USA 103:8113–8118 (2006).

Benton MJ, Donoghue PC: Paleontological evidence to date the tree of life. Mol Biol Evol 24: 26–53 (2007).

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA: Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180 (1981).

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

Brüggerhoff K, Zakhartchenko V, Wenigerkind H, Reichenbach HD, Prelle K, et al: Bovine somatic cell nuclear transfer using recipient oocytes recovered by ovum pick-up: effect of maternal lineage of oocyte donors. Biol Reprod 66:367–373 (2002).

Dollin AE, Murray JD, Gillies CB: Synaptonemal complex analysis of hybrid cattle. II. *Bos indicus* x *Bos taurus* F1 and backcross hybrids. Genome 34:220–227 (1991a).

Dollin AE, Murray JD, Gillies CB: Synaptonemal complex analysis of hybrid cattle. III. Meiotic pairing mechanisms in F1 Brahman × Hereford hybrids. Genome 34:228–235 (1991b).

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

Epstein H, Mason IL: Cattle, in Mason JL (ed): Evolution of Domesticated Animals, pp 6–27 (Longman Group Limited, Essex 1984).

Goldammer T, Brunner RM, Schwerin M: Comparative analysis of Y chromosome structure in *Bos taurus* and *B. indicus* by FISH using region-specific, microdissected, and locus-specific DNA probes. Cytogenet Cell Genet 77:238–241 (1997).

Götherström A, Anderung C, Hellborg L, Elburg R, Smith C, et al: Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. Proc Biol Sci 272:2345–2350 (2005).

Gray AP: Mammalian Hybrids: A Check-List with Bibliography. 2nd ed. (Commonwealth Agricultural Bureau, Farnham Royal, Buckinghamshire 1972).

Hiendleder S: Detailed restriction map of sheep mitochondrial DNA. Arch Anim Breed 36:511–517 (1993).

Hiendleder S: A low rate of replacement substitutions in two major *Ovis aries* mitochondrial genomes. Anim Genet 29:116–122 (1998).

Hiendleder S, Lewalski H, Wassmuth R, Janke A: The complete mitochondrial DNA sequence of the domestic sheep (*Ovis aries*) and comparison with the other major ovine haplotype. J Mol Evol 47:441–448 (1998).

Hiendleder S, Schmutz SM, Erhardt G, Green RD, Plante Y: Transmitochondrial differences and varying levels of heteroplasmy in nuclear transfer cloned cattle. Mol Reprod Dev 54:24–31 (1999).

Hiendleder S, Kaupe B, Wassmuth R, Janke A: Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. Proc Biol Sci 269:893–904 (2002).

Hiendleder S, Zakhartchenko V, Wenigerkind H, Reichenbach HD, Brüggerhoff K, et al: Heteroplasmy in bovine fetuses produced by intra- and inter-subspecific somatic cell nuclear transfer: neutral segregation of nuclear donor mitochondrial DNA in various tissues and evidence for recipient cow mitochondria in fetal blood. Biol Reprod 68:159–166 (2003).

Hiendleder S, Zakhartchenko V, Wolf E: Mitochondria and the success of somatic cell nuclear transfer cloning: from nuclear-mitochondrial interactions to mitochondrial complementation and mitochondrial DNA recombination. Reprod Fertil Dev 17:69–82 (2005).

Jobb G, von Haeseler A, Strimmer K: TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 4: 18 (2004).

Kijas JM, Andersson L: A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. J Mol Evol 52: 302–308 (2001).

Kikkawa Y, Takada T, Sutopo, Nomura K, Namikawa T, et al: Phylogenies using mtDNA and *SRY* provide evidence for male-mediated introgression in Asian domestic cattle. Anim Genet 34:96–101 (2003).

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

Lanave C, Preparata G, Saccone C, Serio G: A new method for calculating evolutionary substitution rates. J Mol Evol 20:86–93 (1984).

Lenstra JA, Bradley DG: Systematics and phylogeny of cattle, in Fries A, Ruvinsky A (eds): The Genetics of Cattle, pp 1–14 (CAB International, Oxon 1999).

Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P: Evidence for two independent domestications of cattle. Proc Natl Acad Sci USA 91:2757–2761 (1994).

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

Mannen H, Morimoto ML, Oyamat K, Mukai F, Tsuji S: Identification of mitochondrial DNA substitutions related to meat quality in Japanese Black cattle. J Anim Sci 81:68–73 (2003).

Mannen H, Kohno M, Nagata Y, Tsuji S, Bradley DG, et al: Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle. Mol Phylogenet Evol 32: 539–544 (2004).

Meadows JR, Cemal I, Karaca O, Gootwine E, Kijas JW: Five ovine mitochondrial lineages identified from sheep breeds of the near East. Genetics 175:1371–1379 (2007).

Nijman IJ, Bradley DG, Hanotte O, Otsen M, Lenstra JA: Satellite DNA polymorphisms and AFLP correlate with *Bos indicus-taurus* hybridization. Anim Genet 30:265–273 (1999).

Nijman IJ, Otsen M, Verkaar EL, de Ruijter C, Hanekamp E, et al: Hybridization of banteng (*Bos javanicus*) and zebu (*Bos indicus*) revealed by mitochondrial DNA, satellite DNA, AFLP and microsatellites. Heredity 90:10–16 (2003).

Posada D, Crandall KA: Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817–818 (1998).

Potter WL, Upton PC: Y chromosome morphology of cattle. Aust Vet J 55:539–541 (1979).

Roubertoux PL, Sluyter F, Carlier M, Marcet B, Maarouf-Veray F, et al: Mitochondrial DNA modifies cognition in interaction with the nuclear genome and age in mice. Nat Genet 35: 65–69 (2003).

Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning, a Laboratory Manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989).

Schmidt HA, Strimmer K, Vingron M, von Haeseler A: TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504 (2002).

Strimmer K, von Haeseler A: Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol 13: 964–969 (1996).

Switonski M, Stranzinger G: Studies of synaptonemal complexes in farm mammals – a review. J Hered 89:473–480 (1998).

Switonski M, Ansari HA, Mathew A, Jung HR, Stranzinger G: Synaptonemal complex analysis in primary spermatocytes of cattle × zebu hybrids (*Bos taurus* × *Bos indicus*). J Anim Breed Genet 107:229–238 (1990).

Swofford DL: PAUP* Phylogenetic Analysis using Parsimony (*and Other Methods), version 4.0b8. (Sinauer Associates, Sunderland 2002).

Thrift FA, Aaron DK: The crossbred sire: experimental results for cattle. J Anim Sci 65:128–135 (1987).

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

van Tuinen M, Hadly EA: Calibration and error in placental molecular clocks: a conservative approach using the cetartiodactyl fossil record. J Hered 95:200–208 (2004).

Verkaar EL, Vervaecke H, Roden C, Romero Mendoza L, Barwegen MW, et al: Paternally inherited markers in bovine hybrid populations. Heredity 91:565–569 (2003).

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

Yang Z: Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39: 306–314 (1994).