Cytochrome b sequences of ancient cattle and wild ox support phylogenetic complexity in the ancient and modern bovine populations

F. Stock*, C. J. Edwards*, R. Bollongino†, E. K. Finlay*, J. Burger† and D. G. Bradley*

*Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland. †Palaeogenetics Group, Institut für Anthropologie, Johannes Gutenberg-Universität, Saarstrasse 21, D-55099 Mainz, Germany

Summary

Mitochondrial DNA has been the traditional marker for the study of animal domestication, as its high mutation rate allows for the accumulation of molecular diversity within the time frame of domestic history. Additionally, it is exclusively maternally inherited and haplotypes become part of the domestic gene pool via actual capture of a female animal rather than by interbreeding with wild populations. Initial studies of British aurochs identified a haplogroup, designated P, which was found to be highly divergent from all known domestic haplotypes over the most variable portion of the D-loop. Additional analysis of a large and geographically representative sample of aurochs from northern and central Europe found an additional, separate aurochs haplotype, E. Until recently, the European aurochs appeared to have no matrilinear descendants among the publicly available modern cattle control regions sequenced; if aurochs mtDNA was incorporated into the domestic population, aurochs either formed a very small proportion of modern diversity or had been subsequently lost. However, a haplogroup P sequence has recently been found in a modern sample, along with a new divergent haplogroup called Q. Here we confirm the outlying status of the novel Q and E haplogroups and the modern P haplogroup sequence as a descendent of European aurochs, by retrieval and analysis of cytochrome b sequence data from twenty ancient wild and domesticated cattle archaeological samples.

Keywords ancient DNA, aurochs, cattle, cytochrome b gene, domestication, mitochondrial haplotypes.

Introduction

All modern cattle were domesticated from the extinct progenitor, the aurochs (Bos primigenius). After the last glacial maximum 15 000 years ago, B. primigenius was found throughout almost the whole of Eurasia and North Africa. While zooarchaeological data points towards the Near East and the Indus Valley as domestication and diffusion centres of modern cattle (Helmer et al. 2005), a debate regarding possible contribution of local aurochs populations, for example in Europe, continues (Götherström et al. 2005; Beja-Pereira et al. 2006; Achilli et al. 2008).

The previous analyses of sequence variation in the mitochondrial chromosome control region have shown a strong phylogenetic structuring of bovine mtDNA haplotypes within Eurasia and Africa. The primary feature of this variation is the existence of two families of sequences (alternately of Bos taurus and Bos indicus origin), each of which clusters in a tight phylogeny but is separated from the other family by a long internal branch. This pattern of distinct, divergent clusters is a recurring feature in domesticated Bovini and likely represents limited episodes of capture from a more phylogenetically complex wild population (MacHugh & Bradley 2001; Finlay et al. 2007).

All B. taurus individuals belong to the T haplogroup, which can be further divided into haplotypes designated as T, T1, T2, T3 and T4, all defined by polymorphisms within 240 bp of the mtDNA D-loop (Loftus et al. 1994; Bradley et al. 1996; Mannen et al. 1998; Cymbron et al. 1999; Troy et al. 2001; Magee et al. 2002). These haplotypes are geographically distributed: T1 is almost exclusively African
(Troy et al. 2001); T, T1, T2 and T3 are all found in the Near East, with T3 predominating in Europe; and haplogroup T4 has been thus far only been detected in Japanese cattle (Mannen et al. 1998). Achilli et al. (2008) also suggest a further taurine haplotype, T5, which is distinguished by variants elsewhere in the mtDNA genome. Bos indicus mtDNA sequences are highly divergent from Bos taurus, and cluster into two haplogroups, I1 and I2 (Baig et al. 2005; Lei et al. 2006; Mugee et al. 2007).

There is an absence in modern samples of intermediate haplotypes, and the predominant phylogenetic branch separating B. indicus and B. taurus is unpopulated. An interesting question is whether the missing phylogenetic complexity of the wild population may be discoverable by either ancient DNA analysis of wild ox or by more extensive sampling of modern cattle. The first study on archaeological aurochs samples was published by Bailey et al. (1996). The authors succeeded in sequencing a 220-bp fragment of the mitochondrial control region from two British B. primigenius individuals. They demonstrated that these individuals, which cluster closely together and were designated as haplogroup P, were divergent from modern cattle and yet positioned more closely to extant taurine than indicine individuals. Haplogroup P is distinguished from modern cattle haplotypes by eight unique control region mutations (Bailey et al. 1996; Troy et al. 2001). From an extensive analysis of control region variation in archaeological skeletal samples ranging from Great Britain to Hungary, it is clearly identifiable as the B. primigenius haplotype of Central, Northern and Western Europe (Edwards et al. 2007). Interestingly, this study also found an additional divergent aurochs haplotype, E. This was discovered in one individual found at the Early Neolithic German site of Eilsleben. This novel D-loop haplotype shares two of the eight mutations that characterize the P haplogroup, but also contains two private mutations and two that it shares with B. indicus and Bison bison individuals. Additionally, it exhibits five mutations that are shared with only very few taurine individuals (Edwards et al. 2007).

Achilli et al. (2008), in an investigation of whole mtDNA genome diversity in modern Near Eastern and European cattle, discovered a third divergent haplotype, designated Q, although this divergence is not marked within the control region alone.

In several surveys of modern mtDNA control region diversity, featuring in excess of 3000 samples from throughout the range of B. taurus, no European aurochs haplotypes had been encountered (e.g. Loftus et al. 1994; Bradley et al. 1996; Mannen et al. 1998; Cymbron et al. 1999; Troy et al. 2001; Magee et al. 2002; Mannen et al. 2004; Hiendleder et al. 2008). However, in 2005, as part of a GenBank deposition of whole mtDNA genome sequences from animals of East Asian and European ancestry, Shin & Kim included a sample, DQ124389, which displays a control region sequence that clearly clusters with the ancient aurochs P haplogroup. The remainder of the mitochondrial genome is also unlike typical modern variation and this deposition may represent a single recorded example of matrilineal introgression of typical European aurochs into the domestic gene pool (Achilli et al. 2008). Shortly before the submission of the whole mitochondrial genome DQ124389 (Shin H.D & Kim L.H., unpublished data) in 2004, Kim S.J., Suh J.Y., Kim K.C. & Suh D.S. (unpublished data) found AY337527, a mitochondrial control region that exhibits the P haplotype. In 2006, Zeng Y.T., Yan J.B. & Huang S.Z. (unpublished data) found sequence data from the mitochondrial control region of a Chinese Holstein (AY998840), appearing to be another living member of the P haplogroup.

Thus, in addition to the two major extant clades of T and I, some sequence information from three other divergent aurochs haplogroups (P, E and Q) are known. Here, we extend the mtDNA sequence haplotypes of 20 ancient cattle samples, including individuals belonging to the control region T, P and E haplogroups. We aligned the cytochrome b gene sequences of 169 published complete mitochondrial genomes and screened for possible clade-specific sites (data not shown). This alignment suggested that DQ124389 (putative modern haplogroup P) differed in two positions from all other B. taurus sequences. At position 14873, DQ124389 exhibited an A whereas all B. taurus, and the majority of B. indicus, individuals showed a G allele. At position 15134, DQ124389 showed a T allele, which it shares with B. indicus individuals. We sequenced 227 bp of cytochrome b, including these informative sites, and used these data to confirm ancient sequence affinity with the single extant P chromosome, and also illustrate the phylogenetic outlying character of the German aurochs E haplotype.

Materials and methods

Samples

A total of 20 Mesolithic and Neolithic aurochs and domestic cattle from five sites in the United Kingdom, Germany and Slovakia were sequenced. All samples had been previously differentiated as domestic or wild, on the basis of size or date, by the researchers who carried out archaeozoological studies of the various sites. All samples included in this study had been genetically assessed and replicated in the previous studies and had been proven to contain sufficient mitochondrial DNA (Bollongino et al. 2006; Edwards et al. 2007). Thirteen of the 20 were classified as B. primigenius; 12 of these (CPC03, CPC05, CPC06, CPC07, CPC08, CPC11, CPC12, CPC13 CPC14, CPC98, EIL6 and NORF) exhibited the ancestral P haplotype of the European aurochs, while one (EIL4) exhibited the novel ancestral E haplotype. Of the remaining seven, five individuals could not clearly be identified as either B. primigenius or B. taurus in the morphometric assessment. Four (CPC04, CPC10, CPC15, and CPC17) had been found to be intermediate between the two species in size and shape.
WH06 and WH10) proved to exhibit the domestic T3 haplotype and only one (SVO3) of them showed the ancestral P haplotype. Two individuals (EIL2 and SVO1) were classified as domestic cattle and showed the T3 haplotype. For further details, see Table S1.

Extraction and polymerase chain reaction

The samples were analysed in two different laboratories: The Smurfit Institute of Genetics at Trinity College Dublin, and the Institute of Anthropology at the University of Mainz (see Table S1). Bone samples were prepared using previously described protocols (Edwards et al. 2007). All primers were designed to be genus specific (Edwards 2002; Bollongino 2005). Polymerase chain reaction (PCR) set-up was carried out in laboratories that were solely dedicated to pre-amplification ancient DNA work. Primer details and PCR cycling conditions are described in Tables S2 & S3. All PCRs were carried out in laboratories that were solely dedicated to pre-amplification ancient DNA work. Primer details and PCR cycling conditions are described in Tables S2 & S3. All PCRs were set up in 50 µl reaction volumes containing one volume PCR buffer [50 mM KCl; 20 mM Tris–HCl (pH 8.4)], 2.5 mM MgCl2, 200 µm dNTPs, 2.5 U PLATINUM® Taq polymerase (GIBCOBRL®) and 4 µmol of each primer.

Statistical and phylogenetic analyses

Ancient DNA is affected by post-mortem alterations of bases, which may result in incorrect sequence determination. In addition, ancient DNA is very prone to contamination. For these reasons, strict criteria for the authentication of mitochondrial DNA sequences were employed. As described elsewhere (Edwards et al. 2004; Bollongino et al. 2006). Five of the 20 samples had previously been replicated in two different laboratories (Edwards et al. 2007). The sequences were manually aligned and a median-joining network was drawn (Bandelt et al. 1995). To put the sequences in a richer phylogenetic context, only those sequences with alignable mitochondrial control region data were included. The ratio of transversions to transitions was established from 1000 pseudoreplicates of the data.

Results

A total of 227 bp of the cytochrome b gene of 20 ancient domestic cattle and aurochs individuals were successfully amplified from bone extractions and sequenced. These
227 bp comprised two fragments from the centre of the cytochrome b gene, ranging from positions 14811 to 14900 and 15031 to 15167 (numbering according to Anderson Reference Sequence; Anderson et al. 1982). One specimen (WH06) could not be unambiguously typed due to low preservation quality and was removed from further analyses. The sequences reported in this study have been deposited in the GenBank database under accession numbers FJ392894–FJ392913.

Within the NCBI database, there are 304 B. taurus and B. indicus complete cytochrome b sequences available. Additionally, two Bison bison individuals were employed as an outgroup. The publicly available sequences included a dataset of 136 sequences sampled from Chinese endogenous cattle breeds (Cai et al. 2007; downloaded May and August 2008). This was the largest contiguous data set and, because of unusual variability, was considered separately. Several other anomalous aspects of these data were apparent. Firstly, this collection of sequences seemed to have a high number of haplotypes (23 in 136 samples) compared with the rest of the bovine sample (12 haplotypes observed in 168 individuals). Secondly, the dataset displayed a significantly higher number of transversions relative to transitions (tv:ti); that is, 8:18 compared with 0:17 in the wider data set (Fisher Exact Test, \( P = 0.011 \)). Note that mtDNA is known to display an especially high transition–transversion ratio and published samples of other Bovini cytochrome b data (banteng, mithun and yak) also show zero or very few transversions despite appreciable transition variation (Anderson et al. 1982). Thirdly, a comparison with the other cattle and Bovini data sets shows the Chinese samples display a much higher non-synonymous to synonymous (Pn/Ps) ratio, with 10:16 compared with 3:14 in other cattle (Fisher Exact Test, \( P = 0.13 \)). Because of these underlying statistics, we decided to exclude this Chinese dataset from further analyses.

The remaining 165 partial cytochrome b sequences, the 19 ancient sequences and the two Bison bison outgroup sequences were aligned (Table 1). The 227-bp fragment displayed 27 polymorphic sites, 10 of which were exclusive to the Bison outgroup. Within the remaining 17 polymorphic sites were eight singleton transitions (14847, 15047, 15068, 15082, 15092, 15102, 15155 and 15157), and six transitions which appear to indicate a subspecies-specific split between B. indicus and B. taurus (14825, 14858, 14897, 15105, 15134 and 15146).

Two of the polymorphic sites, 14873 and 15134, are of particular interest for this study because they seem to differentiate between the B. taurus and the B. primigenius clades. At site 15134, the data suggest that the ancestral allele is a T and the vast majority of domesticated cattle display a C. The only exceptions are the Korean modern P sequence (DQ124389) and the divergent Q haplotype (EU177866 and EU177867). The ancient specimen EIL4 and B. indicus individuals possess the ancestral allele. The sequences gained from the Neolithic samples fall in both groups. The individuals EIL2, CPC04, CPC10, SV01 and WH10 show a C whereas all other sequences exhibit a T at this site. These also possess a T3 D-loop haplotype. Of the remaining 14 ancient samples, 13 (CPC03, CPC05, CPC06, CPC07, CPC08, CPC11, CPC12, CPC13, CPC14, CPC98, EIL6, NORF and SV03) possess D-loop sequences belonging to the aurochs haplogroup P. The individual EIL4 typed with a divergent D-loop haplotype, designated E (Edwards et al. 2007), showed the T allele, supporting its placement as an outlying haplotype. The potentially ancestral allele at site 14873-A is exhibited by 10 of the 14 members of the P haplogroup, including DQ124389. Only two B. indicus and Bison bison individuals share this allele. CPC04 and CPC10 are interesting samples because these bovine bones were robust and fell within a size range typical of aurochs rather than domesticates. However, they gave T3 D-loop haplotypes (Edwards et al. 2007). Here, they also possess the alleles at sites 15134 and 14873 typical of domesticates.

To place the ancient cytochrome b variation in a richer phylogenetic context, we constructed a median network (Bandelt et al. 1995) of the 19 ancient samples analysed here, plus 140 modern cattle samples for which alignable positions in both the cytochrome b gene and control region were available (Fig. 1). Mutations in the D-loop fragment were down weighted in the network from 10 to 1 because of the known presence of hypermutable sites.

The network shows clearly the genetically distinguishable bovine sub-species, B. taurus and B. indicus, separated by six cytochrome b substitutions. All B. indicus haplotypes cluster closely together, as do all B. taurus sequences. All B. primigenius individuals typed with a P haplotype sit closer to the B. taurus cluster than to the B. indicus cluster. Within this P grouping is the modern sample, DQ124389, distinguished from other modern B. taurus at both the 14873 and 15134 sites. At site 15134, the derived state is possessed by each of the T haplogroups, and among modern B. taurus is absent only in the single modern P and the two individuals reported with the divergent Q haplotype; these are separated by four mutations from the T3 cluster.

The unusual haplotype E sequence, EIL4 from a German B. primigenius sample, is distinguished by several cytochrome b mutations, confirming it as an outlier to both modern cattle and other aurochs. However, a neighbour-joining tree of the cattle sequences and a reasonable sample of Bovini outgroups places it within the same clade as B. taurus (Fig. 2). Despite its unusual divergence from modern samples, there is adequate bootstrap support for excluding the possibility that the E bone sample was a different species that was misidentified as aurochs.

**Discussion**

The discovery of the P haplogroup in North and Central European wild oxen supports an important archaeological
premise; that European cattle matrilineal have a domestic
domestic origin in the Near East rather than from local European
strains (Troy et al. 2001). Edwards et al. (2007) examined
59 bone samples from Mesolithic, Neolithic and Bronze Age
European contexts that had been designated as aurochs
by measurement or dating evidence. They found a
mixture of P and T haplotypes; a distinction which we
confirm here in the same samples by further sequencing of
informative regions of the cytochrome b gene. The presence
of the P haplogroup in Central and Northern European
samples confirmed that this aurochs variant is the typical
mainland European type; it had previously only been
encountered in Britain (Bailey et al. 1996; Troy et al. 2001).
The bones giving T haplotypes were either of intermediate
size between wild and domestic or were fragmentary and so
could not be measured effectively and so domestic origin
could not be excluded for these animals. Each amplifiable
archaeological specimen unambiguously identified as
aurochs, either by measurement or time depth, gave a P
haplotype.

In over 10 years of studies of bovine control region
sequence variation in modern cattle from disparate
geographical origins, only three of over 3000 submitted
haplotypes in GenBank show a P sequence: AY337527,
AY998840 and DQ124389. DQ124389 is a complete
mitochondrial genome and has been shown to be distinct
from all domesticates (Achilli et al. 2008). This surprising
finding required confirmation, and thus we here present

**Figure 1** A median phylogenetic network including 159 individuals and a total of 429 bp from both the cytochrome b gene and the D-loop. Colours as follows: T: dark blue; T1(a): yellow; T2: bright green; T3(a): red; T4: grey; T5: pink; Q: black; P: light brown; E: light blue; I1: dark purple; I2: light purple. All cytochrome b mutations are clearly indicated. The size of a circle in the network is proportional to the number of individuals, and the length of the branches is proportional of the number of mutations. An exception to this rule is indicated in the branches leading to haplogroups E, I1 and I2.

**Figure 2** Neighbour-joining tree. Bootstrap values were calculated from 1000 pseudoreplicates of the data. The three highest bootstrap values are given, confirming E as an outlier to the taurine cluster, and thus excluding a misidentification with a member of one of the other Bovini outgroups.
cytochrome b sequence from 13 aurochs P individuals and compare them with DQ124389. Our data confirm that this published modern P individual matches the consensus ancient Northern European aurochs sequence at two diagnostic polymorphic sites, strongly suggesting that this individual is indeed the product of a rare, and seemingly unlikely, matrilineal introgression from the wild into the European domestic gene pool. It is also conceivable that the P lineage originates from the as yet unexplored ancient Asian population. Whereas AY998840 and DQ124389 are sampled from individuals of the Holstein breed, AY337527 is an individual of the indigenous Korean Youngju Yellow breed. Introgression of local Asian aurochs populations has also been suggested by Mannen et al. (1998, 2004) who reported a new taurine haplotype, T4, which has been so far only detected in Asian cattle populations. However, the phylogenetic relationship of T4 to other T haplotypes (most especially the fact that it roots through T; Achilli et al. 2008) does not rule out a Near Eastern/Anatolian origin of all T haplotypes.

The results here confirm some assertions of the survey by Edwards et al. (2007). Importantly, the divergent D-loop sequence of a single aurochs from Eisleben (ELA: haplogroup E) is confirmed as an outlier by cytochrome b sequence data. Edwards et al. (2007) report this as an unusually robust archaeological bone and their data place it phylogenetically within cattle sequences (thus excluding an identity as a Bison or more distant bovid); the results here confirm it as a more divergent relative to modern cattle than haplogroup P. The cytochrome b mutation at site 14873 also suggests the existence of a third rare aurochs sequence type, Q, distinct from modern cattle sequence types, which has already been identified as an outlying sequence by Achilli et al. (2008). Q may be a Southern European or Near Eastern variant, but its provenance is impossible to ascertain as it has been encountered only once, in a double sample from a single Italian breed.

Thus it seems that the missing history in the standard modern cattle phylogenetic structure is populated by rare divergent mitochondrial types. These appear to represent matrilineal captures from the wild, or, alternatively, might point to a higher mitochondrial variability of domestic cattle than previously thought. The genetic variation of B. primigenius was geographically structured – P seems to be the typical European variant, with predominance across Europe, from Britain to the Hungarian plain. Outliers appear rarely, for example E as a single, and morphometrically robust, sample. There is an inadequate sample of aurochs from the Near East with which to build an estimate of phylogenetic complexity, but one might surmise that T-like variation was typical, given its predominance in modern B. taurus. It is very unlikely that the divergent modern P haplotype was domesticated in the Fertile Crescent: its extreme rarity and the provenance of the wild haplotype suggests it to have resulted from a secondary, possibly European, introgression from the wild. Thus, in the course of European domestic history, it does appear that at least one wild-born female cow was captured and bred as a domesticate, an event which these data suggest to be of vanishing rarity.

Acknowledgements
We thank Petra Czerwinski of the Johannes Gutenberg-Universität, Mainz, for assistance with laboratory work. We thank Andrew Chamberlain, Hans Jürgen Döhle, Hans-Peter Uerpmann and Rick Schulting for providing the archaeological bone samples for this study. F.S. is funded by the Genetime Program, a Marie Curie Training Site.

References


Supporting information

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

Table S1 Table of archaeological bone samples studied, with associated codes and skeletal element used.

Table S2 Primer pair sequences, given in 5¢–3¢ direction, and their positions in the mitochondrial genome numbered according to Anderson Reference Sequence V00654 (Anderson et al. 1982).

Table S3 PCR cycling conditions for (a) Dublin and (b) Mainz primer pairs.

Table S4 Sequence information.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors.