Interrogation of modern and ancient genomes reveals the complex domestic history of cattle

David A. Magee,* David E. MacHugh,*† and Ceiridwen J. Edwards‡

*Animal Genomics Laboratory, School of Agriculture and Food Science, College of Life Sciences, University College Dublin, Belfield, Dublin 4, Ireland
*†UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin 4, Ireland
‡Research Laboratory for Archaeology, Oxford University, Oxford, United Kingdom

Introduction

The development of agriculture at the advent of the Neolithic period, circa 13,000 years before present (YBP), is considered to be the most important cultural innovation in human history. Before this, anatomically modern humans had successfully managed to occupy most habitable and accessible regions of the world as small groups of widely dispersed hunter-gatherers, exploiting a diverse range of wild plant and animal species. The transition in human subsistence from nomadic forager to sedentary agriculturist had a number of profound cultural, economic, and demographic consequences. Through the cultivation of crops and the domestication of large animals, humans could manipulate and manage their natural environment in a completely new way. The concomitant increase in food and other biological resources facilitated a significant increase in human population numbers, leading to the geographic expansion of agriculturalist communities and the eventual emergence of complex urban societies. Consequently, deciphering the anthropological and biological processes associated with plant and animal domestication is key to understanding the origins of modern human society (Diamond, 2002).

Domestication has been defined as “the human creation of a new form of plant or animal—one that is identifiably different from its wild ancestor and extant wild relatives” (Smith, 1995); in terms of animal domestication, this can be viewed as an elaboration of the predator–prey relationships between Homo sapiens and target species. Animal domestication is generally a long-term cumulative process, characterized by gradual changes in the behavior of both human and animal populations over time, which has had selective advantages for both biological partners and has been aptly described as a “covenant of the wild” (Budiansky, 1994). As with plants, there are two key elements associated with the initial domestication of animals: at first, it is most likely that humans segregated a few individuals from wild populations; following this, active intervention in the life cycles of captive populations lead to a behavioral and physiological trajectory toward full domestication. This was not an instantaneous event in which a wild animal was immediately transformed into a domesticate; instead it was a process that required sustained breeding of captured and tamed animals over prolonged periods of time, with the change from wild to domestic forms occurring after several domestic generations since initial capture from the wild (Uerpmann, 1996). Sustained human management of early domestic populations led to an increase in animal numbers and an expansion of species’ geographic range beyond that of...
incided with the advent of agriculture (Diamond, 2002; Bar-Yosef, 2011).

During the periods of environmental and climatic unpredictability that co-
incided with the advent of agriculture (Diamond, 2002; Bar-Yosef, 2011).

Zooarchaeological Investigation of Animal Domestication

Much of what is known about the ‘when’ and ‘where’ of animal domestication has been gleaned from the scientific examination of faunal material retrieved from archaeological sites. These zooarchaeological studies of animal domestication are largely based on osteological analysis and radiocarbon dating of faunal remains at early farming settlements. Geographic centers of animal domestication are those where, over time, a staged, temporal transition from characteristically wild to domestic faunal patterns are evident in response to changes in human behavior (Meadow, 1989). In this regard, many of the biological changes that occur due to domestication are morphological, and these observable changes allow archaeologists to ascertain and catalog evidence for animal domestication at a particular site (Clutton-Brock, 2012).

It is thought that selection for docility and reduced aggression in wild and early domesticates were indirectly associated with gradual, yet distinctive, changes in secondary morphological characters, such as the form and fine structure of bones. These include the shortening of the skull that resulted in the crowding and reduced size of teeth, reductions in brain size, and changes in horn morphology (Bökényi, 1976; Smith, 1995). Changes in morphological markers are generally harder to distinguish in early domesticates; for example, Bökényi (1976) considers that well-defined morphological changes do not occur for about 30 generations (approximately 200 years in the case of cattle). Nevertheless, at least two kinds of morphological change are thought to have occurred relatively early in the domestication process. The first is the appearance of bone pathologies, reflecting the physical trauma (caused, for example, by the use of the domesticate for draught), poor diet, and higher infection rates resulting from enforced confinement; the second is a decrease in size due to selection of more manageable animals and malnutrition (Bökényi, 1976; Smith, 1995).

Additional evidence includes temporal changes in the proportions of age and sex groups, whereby the demography of faunal assemblages produced by human hunting differs from those produced from managed herds. This may suggest a pre-selection of animals for domestication (for example, young female animals may have been more easily captured and tamed than wild males) or differential use of sex groups (for example, young males may have been preferentially slaughtered for food rather than reproductively active females). Finally, ancillary archaeological evidence for animal domestication includes artistic representations (rock art

Box 1. Genetics and Domestication

The principles of Darwinian evolution predict that modern populations share biological (and therefore genetic) continuity with past populations of similar geographical provenance. Therefore, modern patterns of genetic diversity reflect signatures of past demographic events. Previous population expansions, contractions (or “bottlenecks”), migrations (to and from populations with subsequent breeding), amalgamations, and selection events (natural or artificial) leave characteristic signatures on genomes in the form of altered allele frequencies, haplotype structures, and selective sweeps, which are transmitted to succeeding generations. The extent of genetic differences between populations or species is indicative of their relationship since they shared a common ancestor. Therefore, modern genomes can be viewed as “molecular heirlooms” that have been transmitted from past to present generations and from which past population histories can be reconstructed. When an ancestral population splits, the frequencies of alleles represented within each separate group will slowly change over time due to the processes of mutation, random genetic drift, and selection. It is these differences that can be assessed to determine the time depth of the divergence event.

Consequently, studies of molecular genetic variation provide a more sensitive tool with which to assess breed and population relationships than measurements of morphological characters.

Investigations of domestic origins and genetic diversity in modern populations can be viewed as an extension of the approach conceived by Nikolai Vavilov in the 1920s (Vavilov, 1926). He realized that elevated phenotypic (and genetic) diversity within a defined geographic area, together with an overlapping distribution between domestic and pre-domestic forms, represents biological evidence for domestication. Accordingly, modern animals originating from a purported domestication center are expected to display elevated levels of genetic diversity due to retention of more wild ancestral variation compared with recently derived populations where a proportion of ancestral variation has been lost through subsampling during successive phases of migration from these centers. Furthermore, as there is a correlation between time and mutation, older populations are expected to have accumulated more DNA sequence variation over time compared with younger populations; therefore, direct modern descendants of the early domesticates should exhibit increased genetic variation as they have had more time to accumulate new mutations compared with younger derived populations.
and figurines) depicting captured wild and domestic animals (Zeder et al., 2006).

The Domestic Taxa of Cattle

Cattle are the most economically important of domesticated animal species and number more than 1300 million on a global scale, constituting some 800 extant cattle breeds (FAO, 2012). Domestic cattle are classified with binomial Linnean names, following convention and the work of Baker and Manwell (1980) and Grigson (1980): Bos taurus denotes humpless taurine cattle and Bos indicus denotes humped zebu (or indicine) cattle. Taurine cattle predominate in northern and western Africa and almost the whole of the Eurasian landmass, from northwest Europe to Japan, whereas zebu cattle are native to the Indian subcontinent and are also found in the Near and Middle East and most of eastern and southern Africa.

Zebu cattle differ from taurine cattle in a number of significant aspects: the presence of a muscular fatty hump of variable size positioned either on the shoulders (cervico-thoracic) or the neck (cervical); a larger, more pendulous dewlap; drooping ears; a naval flap; and tolerance of arid environments. Zebu cattle also display a lower basal metabolic rate, lower water and nutrient requirements, and are generally more resistant to ticks and intestinal parasites than taurine animals. However, unlike zebu cattle, indigenous African taurine cattle display an inherited tolerance to the symptoms of trypanosomiasis, a wasting disease caused by flagellate protozoans of the genus Trypanosoma and transmitted by tsetse flies (Glossina sp.; Payne, 1991).

Despite differences in morphology and disease tolerance/susceptibility, both domestic taxa possess the same number of chromosomes (2n = 60; Melander, 1959). With regard to the sex chromosomes, there is one form of the X chromosome in cattle, which is submetacentric in both taurine and zebu cattle. In contrast, there are two forms of the Y chromosome, with the typical B. taurus Y chromosome morphology being submetacentric and B. indicus having small acrocentric Y chromosomes. As taurine and zebu cattle can freely interbreed, producing fertile offspring, they are often regarded as subspecies within the Bos genus (Bökönyi, 1997).

The Domestic Origins of Cattle

It is now widely accepted that the two modern domestic cattle taxa were derived from the now-extinct wild ox, or aurochs (Bos primigenius). Bos primigenius evolved during the interglacial period of the mid-Pleistocene about 500,000 YBP (Guinnard, 1999) and are thought to be descendants of B. acutifrons (Pilgrim, 1947). During the Middle and Upper Pleistocene, aurochs were widespread throughout Eurasia, and through differences in horn shape and body size, three continental races have been identified: B. primigenius primigenius (Eurasia), B. p. namadicus (South Asia), and B. p. opisthonomus [North Africa] (Grigson, 1978, 1980). In Eurasia, the range of the aurochs expanded and contracted in response to interglacial and glacial stages during the Ice Age. After the last deglaciation, around 12,000 YBP, aurochs populated much of Eurasia and portions of northern Africa, stretching from Britain to China, with the exception of northern Scandinavia, Ireland, and Siberia.

With a derivation from the Greek word oros, meaning mountain, the word aurochs, translated from German, is both singular and plural and literally means “primeval ox” or “proto-ox.” Aurochs bulls were pronouncedly larger than modern domesticates, up to 2 m at the shoulder, and were often equipped with extremely long horns (Zeuner, 1963). Although the last surviving European aurochs reportedly died in a Polish animal reserve in AD 1627 (Zeuner, 1963), zooarchaeological analysis and contemporary descriptions of this animal suggest that the domestication of wild cattle was a considerable achievement. Indeed, Julius Caesar, in his record of the Gallic Wars, wrote of them: “They are a little below the elephant in size and of the appearance, color, and shape of a bull. Their strength and speed are extraordinary; they spare neither man nor wild beast which they have espied….But not even when taken very young can they be rendered familiar to men and tamed.”

The earliest recorded archaeological evidence for cattle domestication comes from the early Neolithic site of Dja’de el Mughara in the Middle Euphrates Valley in southwest Asia (Helmer et al., 2005). Progressive reductions in the sizes of fossil cattle bones through successive depositional layers have been documented, dating to 10,750–10,250 calibrated (cal) YBP, and these are thought to be representative of the first domestic B. taurus cattle derived from the Eurasian aurochs, B. p. primigenius. There is also morphological evidence for management of taurine cattle in northeastern China at this time (10,660 cal. YBP), which may indicate local domestication of autochthonous aurochs (Zhang et al., 2013).

It is generally accepted that taurine cattle, along with other Near Eastern domesticates, such as sheep, goats, barley, and wheat, spread into Europe via two routes: a Mediterranean route and a Danubian (or Continental) route. Archaeological evidence suggests the dispersal of agriculture in Europe began around 9,000 cal. YBP, with the first European farming settlement located at Argissa Magoula in Greece at 8,600 cal. YBP (Rei-
Box 2. mtDNA as a Marker for Population Genetics

Different classes of DNA markers have been used to reconstruct the ancestry and domestic history of cattle. Mitochondrial DNA (mtDNA) has been extensively used to test hypotheses about the domestication of animals. Mammalian mitochondrial chromosomes are small, circular organelar genomes (~16,300 bp) that display a maternal mode of inheritance, an absence of recombination, and a rapid rate of mutation relative to the nuclear genome. Thus, mtDNA lineages are regarded as clonal, information-rich, singly inherited units (or haplotypes) from which matrilineal evolutionary histories can be reconstructed for domestic populations that otherwise may have complex genetic histories.

Typically, phylogenies of wild ungulate mtDNA show multiple clades and a complex branching structure resulting from a wide geographic sampling of wild variability (Bradley and Magee, 2006). In contrast, mtDNA phylogenies constructed from modern cattle show two major distinct clades, representative of *B. taurus* and *B. indicus* sequences. As domestication results in the attenuated, geographically restricted sampling of wild variation, the two major mtDNA clades found in domestic cattle correspond to the mtDNA sequences sampled from the Eurasian and South Asian aurochs landraces during the domestication process. The extant diversity within the *B. taurus* and *B. indicus* clades also contain the limited amount of diversity that has accrued via mutation within the ~10,000-year time frame of cattle domestication (Figure 1B).

The consequence of domestication for mtDNA sequence diversity is illustrated in the figure below where phylogenetically distinct mtDNA sequences (colored circles, in which similar mtDNA sequences show similar shading) sampled from wild extant populations across wide geographic locales generate “bushy” phylogenies consisting of multiple clades with complex branching structures. In contrast, the attenuated sampling of mtDNA sequence variation from geographically and genetically disparate wild populations causes sequences to cluster into small numbers of clades, often generating “double-headed broomstick” topologies, such as that observed for *B. taurus* and *B. indicus* (MacHugh and Bradley, 2001).

Studies have also shown that for domestic mammalian livestock, mtDNA sequences show geographic inertia due to maternal inheritance and low migration of females under managed breeding (i.e., modern domestic animals sampled from similar locations retain distinct mtDNA haplotypes despite millennia of interbreeding with animals, usually males, imported from other locations). Consequently, sequences that originate from different captures of a diverse wild species can maintain a phylogenetic distinction even after millennia of predominantly male-mediated domestic inbreeding, which allows the resolution of pre-domestic patterns of diversity (for example, the sequence divergence between modern *B. taurus* and *B. indicus*) and post-domestic patterns of sequence diversity (for example, diversity within each of the major taurine and zebu haplogroups) to be discerned. The majority of bovine mtDNA phylogeographic studies have targeted the hypervariable mtDNA control region or displacement loop (D-loop); however, more recent studies have analyzed the complete mitochondrial genome.

![mtDNA phylogeny](image)

**Geographic distribution of wild mtDNA sequences**

**Phylogeny of wild mtDNA sequences**

**Sampling of wild mtDNA sequences from different geographic locales during domestication**

**Phylogeny of domestic mtDNA sequences**
Evolutionary history of male lineages can be examined using polymorphic markers distributed across the paternally inherited Y chromosome. Y chromosome markers largely consist of simple tandem repeats (STR), such as microsatellites, which are tandemly repeated short non-coding DNA sequences (for example, CACACACACACACA). These repetitive arrays are often highly polymorphic in length due to variation in the number of repeat units, which generates a range of alleles that are transmitted in a Mendelian fashion (Underhill and Kivisild, 2007). The STR are ubiquitous and distributed throughout the mammalian nuclear genome and can also be used to analyze genetic variation in the autosomal genome exclusive of the Y chromosome. Generally, autosomal and Y chromosomal STR marker length variation is presumed to be selectively neutral, unless the locus under examination is closely linked to a selected coding region (Rosenbaum and Deinard, 1998).

More recently, whole-genome DNA sequencing has revealed the abundance of SNP that exist in eukaryotic nuclear genomes. These are defined as single base pair substitutions in genomic DNA, where the least common allele has a frequency of ≥0.01 in a particular population (Kruglyak and Nickerson, 2001). On average, SNP occur once every 500 to 1,000 bp of mammalian DNA sequenced and are found in both coding and non-coding nuclear DNA (Bovine Genome Sequencing and Analysis Consortium et al., 2009). Although the vast majority of SNP in mammalian genomes are biallelic (owing to their low mutation rate) and therefore have reduced information content compared with STR, their pan-genomic distribution and digital nature (where a SNP locus is queried for the presence of absence of an allele) has facilitated development of high-throughput genotyping arrays or “SNP chips” (Matukumalli et al., 2009). This has resulted in their widespread use in more recent population genetic investigations in cattle and other domestic livestock.

Although archaeological evidence suggests that the earliest domestic cattle of the African continent were humpless B. taurus, and were present in North Africa, the eastern Sahara, and near the Nubian Nile by 7,550–6,950 cal. YBP (Wendorf and Schild, 2001), their origins remain controversial. Traditionally, it was held that B. taurus was introduced into Africa via the Isthmus of Suez from the Near East approximately 9,000–7,000 YBP (Epstein and Mason, 1984); however, some authors contend that the founding B. taurus of Africa was derived from an independent domestication of the indigenous aurochs, B. p. opisthonomus, in the eastern Sahara 9,500–8,900 cal. YBP (Wendorf and Schild, 1994; Grigson, 2000; Wendorf and Schild, 2001; Marshall and Hildebrand, 2002).

Bos indicus cattle were initially thought to have entered sub-Saharan Africa via the Horn of Africa from southwest Asia between 700 and 1500 AD, before and during Arab migrations (Epstein and Mason, 1984; Ho-Burani, 1991). However, evidence from Ngamuriak, a Neolithic site in Kenya, suggests that these cattle breeds were present in East Africa at least 1,500 years earlier (Marshall, 1989). Thereafter, exogenous zebu and native taurine cattle were crossbred to form populations of hybrid sanga cattle, the modern descendants of which are found in southern and eastern Africa. Internal African pastoralist migrations are believed to have been responsible for dispersals of sanga cattle into southern and central African regions (Epstein and Mason, 1984).

The terms “wild” and “domestic” suggest a clear dichotomy between forms and do not consider the slow gradual transition in forms that occurred as human–animal relationships intensified during the domestication process. In contrast, geneticists investigating animal domestication do so within the context of a clearly defined process: the genetic response to the managed breeding of animals. As genetic information is transmitted to succeeding generations, modern populations carry in their genome the signatures of past demographic processes and the evolutionary forces that have shaped these. Consequently, all modern domesticates have descended, with modification, from the wild animals that were incorporated into a finite genetic pool at various stages within the time frame of domestication. Modern domestic populations, therefore, are at the endpoint of a temporal genetic continuum that stretches from the initial phases of animal domestication and includes only those influential interactions between human and animals that have left a genetic legacy (Box 1).

Geneticists examining the origins and ancestry of domestic animals focus largely on the analysis of evolutionarily neutral, non-coding loci of the nuclear genome, for example, autosomal and Y chromosome simple tandem repeat (STR) and single-nucleotide polymorphisms (SNP), and extranuclear genomes, such as the mitochondrial genome (mtDNA). These alternate genetic markers, which display different modes of inheritance, information content, and rates of mutation, have been used to unravel the domestic ancestry of cattle (Boxes 2 and 3). More recently, high-throughput genomic technologies such as SNP genotyping array platforms and whole-genome sequences from animals of wide geographic provenance have enabled high-resolution analysis of the genetic relationships among breeds and have identified discrete segments of cattle genomes that have been subject to the selection processes that underlie domestication.
Wild-ranging surveys of genetic variation in modern cattle populations have been instrumental in showing that all domesticated cattle derive from either taurine or zebu wild progenitors or are hybrids of the two domestic taxa. The most striking aspect of these studies has been the repeated support for the deep divergence between modern B. taurus samples from Africa and Europe and B. indicus animals from India (Fig. 1A-C). This is illustrated by the high number of nucleotide differences between taurine and zebu mtDNA sequences (Fig. 1A), in which, in turn, give rise to the long internal branch that separates the B. taurus and B. indicus clades in the phylogeny shown in Fig. 1B. The extent of sequence divergence between the taurine and zebu mtDNA sequence groups has been estimated in the order of 200,000 years. These patterns of mtDNA diversity indicate the derivation of B. taurus and B. indicus from at least two geographically separate and genetically divergent wild aurochs populations, presumably the Near Eastern and Asiatic aurochs landraces, respectively, as suggested from zooarchaeological evidence (Lofus et al., 1994a,b).

Notably, genetic diversity within the major taurine and zebu phylogenetic mtDNA sequence groups is markedly (~10 times) smaller than that between them (Achilli et al., 2009). This is illustrated by the small number of nucleotide differences within the taurine and zebu mtDNA sequence groups (Fig. 1A) that generate the short-length branches within each clade (Fig. 1B). As natural populations are expected to display geographic and genetic correlation (such that individuals of similar geographic location displayed increased related genetic variation), the patterns of diversity within each bovine mtDNA sequence clade are consistent with the attenuated sampling (that was limited in time and space) of a subset of aurochs variation concomitant with the domestication process. The extent mtDNA sequence diversity within each clade also includes the limited variation that has accrued through mutation post-domestication (Box 2).

The clear genetic distinction between the taurine and zebu mitochondrial genomes can be used to detect regions where historical gene flow and introgression between the two taxa have occurred since the Neolithic. For example, several studies of Near and Middle Eastern populations have reported zebu mtDNA sequences from samples that are phenotypically and morphologically taurine (Troy et al., 2001; Edwards et al., 2007b; Kantanen et al., 2009). While these populations have appreciable frequencies of zebu mtDNA haplotypes, the quantitative divergence between the taurine and zebu haplotypes remain. This has allowed these phylogeographical exceptions to be securely identified as secondary introductions, which is most likely a legacy of the westward introduction of zebu cattle and subsequent mating with native taurine populations (Troy et al., 2001; Edwards et al., 2007a). With some symmetry, the modern zebu breeds of northern India have been shown to possess a minority component of taurine mtDNA ancestry (Magee et al., 2007; Lenstra et al., 2014). Reciprocal migrations of zebu and taurine cattle from their respective geographic centers of origin, and subsequent mating with native populations elsewhere, were presumably a consequence of the extensive ancient trading routes across southwest Asia, which would have provided an opportunity for genetic exchange between the two taxa (Zeuner, 1963; Naik, 1978; Misra, 2001). In support, available archaeological evidence points towards the presence of zebu cattle in Jordan circa 3,400 YBP and taurine

Figure 1A. Alignment of partial mtDNA control region sequences sampled from modern European taurine; African taurine (N’Dama), sanga (Fulani) and zebu (Butana and Kenana); and Indian zebu populations. Only variable sites are shown. Nucleotide substitutions are given; asterisks (*) denote nucleotide insertion/deletions; dashes (-) denote nucleotide identity. The position of the nucleotide differences within the mtDNA control region are reported less 16,000.

Figure 1B. Neighbor-joining phylogeny constructed from modern B. taurus, B. indicus and extinct B. primigenius mtDNA control-region haplotypes. Analysis is based on a 201 bp fragment for the hypervariable bovine mtDNA control region. The quantitative divergence between modern B. taurus and B. indicus supports the independent origins of the two domestic taxa. The limited diversity within each sequence group is consistent with the attenuated sampling of diverse wild matrilines within each domestication center during the domestication process together with the diversity that has accrued since that time (Box 2). The ancient British B. primigenius sequences cluster tightly and form an outgroup to the modern B. taurus haplotypes, suggesting that there was no major maternal contribution to the European domestic cattle genetic pool from locally recruited wild cattle. Figure adapted from Troy et al. (2001). Figure 1C. Neighbor-joining phylogeny of the major bovine mtDNA haplogroups reported in modern and ancient samples based on complete mtDNA genome sequences (16,338 bp). Modern B. taurus predominantly possess the T1, T2, T3 and T4 haplogroups, while modern B. indicus possess haplogroups I1 and I2. Haplogroup P is the major haplogroup from the extinct Eurasian aurochs. Haplogroups O and R have been recently reported in a small number of modern Italian samples and haplogroup C has been recently reported in an ancient Chinese domestic sample; these data support minor local captures of wild matrilines. Yak mtDNA is included as an outgroup species.
Figure 2. Reduced median network analysis of the most variable 240 bp region of the bovine mtDNA control region shows the regional differences in diversity within modern domestic B. taurus (A) and B. indicus (B) cattle, superimposed on a map of regions of origin. B. taurus and B. indicus sequences predominantly fall into five and two major haplogroups, respectively. The positions of the nucleotide substitutions (+16,000) that distinguish each of the major taurine and zebu haplogroups are shown in the skeleton network in the (upper inset in both diagrams). The haplotypes encountered in each region (shaded circles) and the haplogroup to which these haplotypes belong are denoted by the color scheme. Unsampled intermediate nodes or unsampled primary haplotypes (small points) are also shown. Circle areas are proportional to the number of individuals possessing each mtDNA haplotype. The star-like phylogeny surrounding each of the major taurine and zebu mtDNA centrally-positioned, numerically-predominant and presumed ancestral haplotypes is consistent with past demographic expansions associated with the domestication process. The predominant European haplogroup, T3 (shown in red), is a subset of the diversity encountered in the Near East, where haplogroups T (blue), T2 (green) and T3 are encountered at appreciable frequencies, suggesting that European taurine maternal lineages owe their ancestry to primary domestication centers proximal to the Fertile Crescent rather than to local input from the wild. Both African and Far Eastern populations display two additional clusters, T1 (yellow) and T4 (magenta), which is suggestive of matrilineal input from local wild oxen. In South Asia, zebu sequences fall into either the I1 (grey) or the I2 (brown) haplogroups. Figure adapted from Troy et al. (2001) and Magee et al. (2007)

**Mitochondrial DNA Sequence Diversity Reveals the Complex Ancestry within \textit{B. taurus}**

In-depth surveys of partial and complete mtDNA sequence diversity have uncovered more fine-scale aspects of \textit{B. taurus} ancestry. This is demonstrated by the reduced median phylogenetic network analyses (Bandelt et al., 1995) presented in Figure 2. In these, the nucleotide sequence of the most variable 240 bp portion of the bovine mtDNA control region from each sample is aligned to a reference bovine mtDNA genome sequence. The number and nucleotide position of each nucleotide difference from the reference sequence are cataloged for each sample. The data are then displayed visually: distinct mtDNA sequences (or haplotypes) are represented as circles, the areas of which are proportional to the number of samples sharing the same sequence. Nucleotide substitutions, insertions and deletions (i.e., point mutations) that differentiate mtDNA sequences, are represented by lines; the length of each line is proportional to the number of point mutations. Sequences of mtDNA that cluster with each other (based on sequence similarity) form sequence groups or haplogroups.

Inspection of the reduced median networks in Figure 2 reveals that the majority of modern taurine mtDNA sequences fall into one of five phylogenetically distinct, yet closely related, star-like haplogroups, termed T, T1, T2, T3, and T4. Each haplogroup consists of a numerically predominant, centrally positioned haplotype (also termed T, T1, T2, T3 and T4), through which all derived haplotypes root or “coalesce” (Mannen et al., 1998; Troy et al., 2001). A sixth haplotype, T5, has also been recently reported in a small number of modern Italian animals (Achilli et al., 2008, 2009). The numerical and topological predominance of these main haplotypes suggest that they are ancestral sequences incorporated into the domestic pool as a result of the capture and taming of wild female progenitors during the domestication process and which have survived in modern populations. The associated sub-haplotypes are thought to be derived from the ancestral sequence due to the accumulation of mutations since domestication. The star-like patterns of diversity observed within each haplogroup represent the phylogeny expected from a population that has undergone a domestication-induced demographic expansion from a small base (Figure 2). This model proposes that a relatively small number of wild female animals were incorporated into early domestic flocks, which gave rise to increasing numbers of offspring (and consequently, increased mtDNA lineage survivorship) under continuously improving systems of animal management (Bradley and Magee, 2006; Bollongino and Burger, 2007).

Another noteworthy feature of \textit{B. taurus} mtDNA sequence diversity is the spatial distribution of the major haplogroups (Figure 3). In particular, one haplogroup (T3) predominates in Europe and is one of four haplogroups (T, T1, T2, and T3) detected in the Near East at varying frequencies. Haplogroup T2 has appreciable frequencies in Italian, Balkan, and Asian taurine cattle but is found only sporadically in the taurine cattle of central, western, and northern Europe. Also, taurine mtDNA sequence diversity levels are highest in the Near East with reduced diversity observed outside this region. Vavilov’s principle (Vavilov, 1926) is evident in these observations, in which he predicted more ancestral biological variation would be retained at a center of origin for a domestic species. In modern genetic terms, loci in these areas should exhibit higher levels of haplotypic and nucleotide diversity. Lineages would then be lost through successive colonization events, resulting in reduced diversity in derived populations at a distance from the source. The geographical distribution of taurine mtDNA haplogroups, whereby European diversity is a subsample of the diversity encountered in the Near East, supports the hypothesis that European cattle are a derivative of the Near Eastern Neolithic complex (Troy et al., 2001; Lenstra et al., 2014). This is supported by archaeological evidence and indicates that \textit{B. taurus} domesticates migrated from the Near East with nomadic pastoralists into central and northern Europe following the Danubian route, finally reaching western Europe and Britain around 6,000 cal. YBP (Zilhao, 2001; Brown, 2007).

Haplogroup T1 almost exclusively describes African mtDNA diversity and is found only at low frequencies elsewhere; in Africa, non-T1 haplogroups have been reported only in Egyptian and northwest African populations (Troy et al., 2001; Lenstra et al., 2014). The over-representation of the T1 haplogroup in Africa has been cited as evidence for the separate domestication of the founding populations of the African continent, as previously proposed by archaeological data (Stock and Gifford-Gonzalez, 2013). The near-exclusive predominance of the T1 haplogroup in Africa has enabled geographical exceptions to be identified as secondary migrations from Africa. For example, the detection of African T1 haplogroups in Iberia and Sicily and southern Italy suggests African genetic influence into southern Europe via the Mediterranean Sea coastline (Cymbor et al., 1999; Anderung et al., 2005; Beja-Pereira et al., 2006).

T4 is a subgroup of the T3 haplogroup that has been found in the taurine populations of East Asia (including eastern Siberian, Chinese, Korean, and Japanese cattle) but has not been reported in western Eurasian populations (Figure 2). This has led to some speculation for either another possible domestication center in the Far East or the recruitment of wild matrilines into early domestic herds migrating eastwards from the Near East (Mannen et al., 1998; Lai et al., 2006; Kantenan et al., 2009).

Notably, the quantitative difference between the five major taurine ancestral haplotypes is small, with a coalescence time around the central T haplotype estimated at ~16,000 YBP (Troy et al., 2001; Achilli et al., 2008, 2009; Ho et al., 2008; Achilli et al., 2009). Thus, the shallow mtDNA sequence divergence observed in modern taurine populations suggests that the population of wild aurochs from which \textit{B. taurus} was derived was itself limited in diversity.

**Mitochondrial DNA Sequence Diversity Supports a South Asian Derived Origin of \textit{B. indicus}**

A complex picture of \textit{B. indicus} domestication has also emerged from mtDNA studies. Similarly to taurine mtDNA, zebu haplotype sequences fall into two major haplogroups (I1 and I2), comprised of two ancestral haplotypes (also called I1 and I2), through which all other sequences in the phylogeny root (Figures 1C and 2). Comparable to taurine mtDNA variation, the dual star cluster motif within these clusters suggests that they are each signatures of domestication-induced population expansion; the estimated time-depths of the expansions are compatible with zebu domestic history as determined from archaeological studies (Ho et al., 2008). The I1 and I2 haplogroups display some geographic distribution across south Asia, with haplogroup I1 predominating in cattle that have moved eastwards to Southeast Asia from the Indus Valley, the most likely center of origin for this haplogroup. Haplogroup I2 is encountered almost
exclusively within the Indian subcontinent and only sporadically outside this region. Current data support northern India as the center for origin for this haplogroup, which expanded sometime after the expansion of I1. The current distribution of I2 across the Indian subcontinent is most readily explained by later diffusion by trade throughout the region (Baig et al., 2005; Magee et al., 2007; Chen et al., 2010). That modern zebu mtDNA sequence variation comprises two phylogenetically distinct, yet closely related, haplogroups reflects both the limited ancestral mtDNA variation incorporated into zebu domestic pools during the domestication process (and which has survived in modern populations) and the limited variation that has accumulated since this time through mutation.

**Nuclear Genetic Data Support the Independent Domestic Origins of B. taurus and B. indicus and the Mosaicism of African Cattle**

A major drawback of mtDNA-based studies is that they focus only on a single segregating locus with a uniquely maternal mode of inheritance and do not, therefore, represent ancestry across the whole genome. For example, the effects of male-mediated gene flow, which also underlies modern animal breeding and selection strategies, cannot be discerned through the analysis of this exclusive matrilineal genetic marker. Therefore, it is important that the information gleaned from mtDNA is complemented with analyses of autosomal and Y chromosomal polymorphisms.

Analyses of autosomal and Y chromosomal genetic markers (Box 3) also show the marked divergence between the two domestic taxa, as discerned by mtDNA analysis. In particular, these nuclear studies have revealed that alternate collections of alleles from several nuclear loci display high frequencies in either modern taurine and zebu populations, respectively, providing further evidence of separate domestic origins of the two major domestic types and the pronounced time depth in their genetic separation (MacHugh et al., 1997; Hanotte et al., 2002).

Nuclear markers have also been instrumental in teasing out the historical migratory routes of cattle from their centers of origin across the Old World as a consequence of early pastoralist diasporas. This is particularly evident in Africa, where analysis of nuclear genetic markers has shown that modern African zebu populations possess autosomal genomes more similar to Indian *B. indicus* despite all sampled modern African cattle having mtDNA sequences that are clearly *B. taurus* in origin (MacHugh et al., 1997; Hanotte et al., 2002; Bovine HapMap Consortium, 2009; Lenstra et al., 2014). This finding is further supported by the use of autosomal and Y chromosomal zebu-specific markers (that is, markers that are present at high frequencies in Indian *B. indicus* populations compared with *B. taurus* populations), where a cline of east–west zebu-derived alleles running in an east–west direction across the continent and a north–south direction in West Africa is evident (Bradley et al., 1994; MacHugh et al., 1997; Free...
The study of ancient DNA is a sub-discipline of biomolecular archaeology, which encompasses extraction of any biomolecule (such as DNA, blood proteins and collagen fibers, resins, fats, and oils) from ancient tissues. The analysis of DNA from archaeological and palaeontological samples has the potential to answer questions about the past directly, thus providing a temporal aspect to modern genetic studies. Ancient studies have shown that DNA can survive intact for more than 500,000 years, enabling many evolutionary questions to be addressed over this time period, including biogeography, phylogeny, taxonomy and selective processes at the population level, as well as domestication origins. Ancient DNA can be extracted from a variety of animal remains, ranging in condition and age, the most popular of which are bones, teeth, and hair. The cytoplasmic location and high copy number of mitochondrial chromosomes within each eukaryotic cell (estimated at up to 10^3 molecules per cell) has made mtDNA the primary genetic marker for ancient DNA studies. However, recent advances in genomic technologies have seen an increase in the number of ancient DNA studies that involve partial or whole nuclear genome sequences from archaeological samples (Shapiro and Hofreiter, 2014).

Phylogenomic analyses of high-density SNP genotyping data involving tens of thousands of biallelic loci (Box 3) have enabled fine-scaled population genetic analyses previously undetected from mtDNA and STR data. For example, Gautier et al. (2010) showed that unsupervised statistical analysis of pan-genomic SNP data permitted samples to be unambiguously assigned to their population of origin; thus, illustrating the power of high-density SNP data to assess the fine-scale genetic structure of modern breeds. These results reflect the differences in allele frequencies that exist between breeds, including those of similar geography and with shared coancestry. These genetic differences have been generated over the past 200 years due largely to the subsampling of ancestral variation and evolutionary and reproductive isolation.

More importantly, high-density SNP data have enabled detailed high-resolution phylogenetic analysis from which detailed descriptions of cattle domestic history can be elucidated (Decker et al., 2009; Bovine HapMap Consortium, 2009; Gautier et al., 2010; Decker et al., 2014). As with early studies involving mtDNA and STR loci, the data demonstrate the pre-domestic divergence of taurine and zebu cattle, supporting their separate domestic origins (Figure 4). Migrations (and subsequent hybridization) of animals from domestic centers of origin resulted in the appreciable levels of zebu and taurine admixture present in the morphologically taurine and zebu populations of Asia, respectively. The data further confirm the zebu ancestry of African cattle, ranging from 0–20% in West African taurine populations and increasing to 23–74% in southern, central, and eastern African hybrid populations (Figure 4; Decker et al., 2009, 2014). Zebu ancestry has also been detected in modern Italian populations (but not in northern and central European and Iberian populations), which may reflect the Roman importation of Near Eastern cattle (in which zebu admixture had already occurred) to South Europe (Decker et al., 2014). Furthermore, partial African taurine ancestry of Iberian and Italian cattle was confirmed and is presumably a legacy of historical interactions across the Mediterranean.

Despite African taurine cattle being more divergent from European and Asian taurine populations (Figure 4), pan-genomic SNP analyses show that taurine cattle from all three continents shared a single recent common ancestor, supporting a single Near Eastern domestic origin for B. taurus (Decker et al., 2014). Migrations of taurine domesticates from this center of origin was followed by admixture with Asian and African aurochs with possible minor captures of indigenous aurochs (Decker et al., 2009, 2014). It is also postulated that the Near Eastern-derived mtDNA haplotypes of early African domesticates was replaced with the divergent T1 haplogroup as a result of the hybridization between migrating domestic populations and resident wild African populations (Decker et al., 2014). Purifying natural selection, in which certain wild mtDNA haplogroups conferred an advantage in African environments, may have also accelerated replacement of Near Eastern-derived haplotypes in Africa (Soares et al., 2013). Analysis also supported the introgression of southern European...
cattle into Far East Asian populations, resulting in the introduction of European taurine and African taurine and zebu genetic contributions (a feature of southern European populations) to the region (Decker et al., 2009, 2014). The current model for the recent evolution and history of domestic cattle, which is based on the analysis and interpretation of high-density SNP data, is presented in Figure 5.

**New World Cattle Display Genetic Ancestry from Multiple Independent Locations**

The European discovery and conquest of the Americas prompted the movement of people and cattle to the New World. Domestic cattle did not exist on the American continent until AD 1493, when Christopher Columbus landed a small number of Spanish cattle on the Caribbean Island of Hispanola. Cattle were continually imported to the region over the following 50 years as agro-pastoralism intensified (Wilkins, 1984), and the majority of the cattle transported to the Americas during this time were of Andalusian (southwest Spain) origin. However, some authors contend that cattle from the Canary Islands, which had been populated with breeds from northwest Spain, were also introduced (Primo, 1992). By the mid-16th century AD, cattle of Spanish ancestry populated most Caribbean islands and had reached Central and South America; these animals are the ancestors of the modern *Criollo* breeds. The 16th century AD also witnessed the importation of Portuguese cattle to Brazil, imported either directly from Portugal or from the Cape Verde Islands (de Alba, 1978). In addition, trade links between northern Europe and North America, as well as Australia, established as part of the European colonization of North America in the 17th century, led to the introduction of several European populations, which were subsequently crossed to generate the hybrid breeds that predominate in North America today.

Between the 16th and 18th centuries AD, African *B. taurus* cattle were introduced to the Americas, possibly as a consequence of the Atlantic slave trade and secondary economic activity. It is believed that these cattle were shipped directly from the French, Portuguese, and Spanish colonies located on the Guinean Gulf coast (Maillard et al., 1993; Maillard and Maillard, 1998). Purchases of West African cattle by the later-established
French settlers of Martinique and Guadeloupe during the 17th and 18th century AD have also been reported (Maillard and Maillard, 1998). Towards the beginning of the 19th century AD, importations of zebu cattle became widespread throughout the Americas. Zebu animals were initially imported from Senegal and were employed to supply the traction necessary for crop cultivation (Rouse, 1973). During the 20th century AD, African zebu animals were gradually replaced by Indian zebu animals, which exhibit increased durability in tropical environments. Crossbreeding of zebu cattle with native Creole animals led to marked rises in cattle productivity in tropical regions, and consequently, large-scale importations of Indian zebu cattle became widespread throughout the Americas during this time (Wilkins, 1984).

Mitochondrial DNA sequence analyses of modern Caribbean and Central and South American cattle reflect the multifaceted historical importations of Eurasian cattle to the Americas. These studies have revealed the predominance of European T3 haplogroup sequences in these populations with varying, yet considerable, frequencies of sequences belonging to the T1 haplogroup. Of note, the derivative T1 haplotype, T1c1a1, which is only sporadically encountered in Iberian populations, has been detected at frequencies of 31% in the populations of the Caribbean Lesser Antilles islands and 50% in modern Brazilian Criollo populations (Figure 2; Magee et al., 2002; Carvajal-Carmona et al., 2003; Mirol et al., 2003; Miretti et al., 2004; Ginja et al., 2010).

The numerical predominance of a peripheral T1 derivative haplotype within these populations suggests a founder effect, in which a restricted subset of the African-derived variation within the Iberian parental population has been retained in modern New World populations. However, this finding does not preclude the direct importations of African cattle to the region. Indeed, genetic studies have revealed the presence of zebu-specific STR alleles sans zebu mtDNA sequences, a feature of African zebu populations, within modern Caribbean populations, thus supporting West African zebu cattle importations to the Americas (Magee et al., 2002; Liron et al., 2006). This is further supported by phylogenetic analyses of New World cattle involving STR markers, which demonstrate that African zebu and taurine cattle contributed significant genetic components of modern American cattle (Delgado et al., 2012; Martinez et al., 2012). Recent analysis of high-density SNP data confirm the genetic affinity between New World and Iberian cattle and has revealed levels of African admixture in modern American breeds that is higher than previously estimated from historical records and genetic analyses based on STR and mtDNA data; these results suggest that the African ancestry observed in modern American populations is, in part, of direct African origin (Gautier and Naves, 2011). However, more recent SNP-based studies of American cattle genetic diversity suggests that this African genetic influence was present in Iberian cattle before the introduction of cattle to the New World; these findings support an Iberian origin for the African ancestry observed in modern New World populations (McTavish et al., 2013; Decker et al., 2014).

More recent Indian zebu genetic influence in New World populations is supported by several lines of evidence including mtDNA (Magee et al., 2002; Paneto et al., 2008), Y chromosome microsatellites (Giovambattista et al., 2000), and autosomal STR and SNP genetic markers (Liron et al., 2006; Egito et al., 2007; Achilli et al., 2009; McTavish et al., 2013). Interestingly, McTavish et al. (2013) suggest that the zebu ancestry observed in New World cattle may have originated, in part, from the founding Iberian cattle of the Americas; however, the lack of zebu ancestry in Iberian cattle has led Decker and colleagues (2014) to refute this and propose that zebu introgression in the Americas occurred after the arrival of Iberian cattle in the New World.

The Promise of Ancient DNA for Studies of Cattle Domestication

Due to the high copy number of mitochondrial genomes in the cell compared with the nuclear genome, ancient DNA research in cattle has been concentrated primarily on the analysis of mtDNA variation (Box 4; MacHugh et al., 1999; Troy et al., 2001; Bollongino and Burger, 2007; Edwards et al., 2007b; Stock et al., 2009). Studies have been successful in addressing some broad questions, such as the nature of domestication and the relationship of domestic cattle to the European wild aurochs. Mitochondrial DNA sequences from aurochs skeletal remains sampled from central and northern Europe (including Britain) belong largely to a single haplogroup, P, which is phylogenetically differentiated from modern European B. taurus animals (Figures 1B and 1C; Troy et al., 2001; Edwards et al., 2007b). In addition to haplogroup P, another haplotype, E, was detected in a single specimen from the Early Neolithic site of Eilsleben in Germany (Edwards et al., 2007b).
The absence of European aurochsen mtDNA sequences in modern European *B. taurus* animals supports the hypothesis that autochthonous aurochs were not domesticated within Europe, but instead that domestic cattle were derived primarily from populations domesticated in the Near East (Troy et al., 2001; Edwards et al., 2007b). However, localized aurochs domestication within Europe cannot be completely discounted; Achilli et al. (2009) identified two novel haplogroups (Q and R) in modern Italian cattle that are considerably different to those previously described in modern *B. taurus* populations. Also, a single European *B. primigenius*-defining haplogroup P sequence obtained from a modern animal has been deposited in the GenBank sequence repository (Achilli et al., 2008) while another haplotype (haplogroup C), has been identified in an archaeological assemblage from early Holocene China (Figure 1C; Zhang et al., 2013). Collectively, these data support minor captures of wild matriline during the colonization of Europe. In addition to mtDNA, ancient DNA studies using nuclear markers have aimed to address the long-standing question of the localized domestication of European aurochs or male-mediated introgression from the wild into domestic herds; however, results from these studies have remained equivocal (Götherström et al., 2005; Bollongino et al., 2006; Svensson and Gotherstrom, 2008).

Despite the availability of a large set of modern data, the information content of cattle mitochondrial haplotypes limits their use for ancient DNA studies; while valuable for global phylogeography, current western European cattle populations display little mitochondrial population genetic structure (Lenstra et al., 2014). On the other hand, nuclear DNA data have been shown to provide a breed-level genetic resolution within modern cattle and thus a more nuanced picture of genetic variation across Europe (Bovine HapMap Consortium, 2009; Gautier et al., 2010; Edwards et al., 2011). Therefore, where possible, extension of nuclear DNA-based studies from extant populations to archaeological specimens will greatly help our understanding of ancient population affinities and local domestication events.

**New Genomic Technologies and the Future of Cattle Domestication Studies**

The availability of a well-annotated taurine bovine genome sequence (Bovine Genome Sequencing and Analysis Consortium et al., 2009), with associated advances in high-throughput genomic technologies, have revolutionized genomic studies and immediately suggests powerful new approaches to better understand cattle domestication. Furthermore, high-throughput resequencing projects using so-called “next-generation” technologies, whereby millions of genomic DNA fragments from an individual are randomly generated, sequenced in parallel and analyzed, have generated whole-genome sequences for several animals, including animals of economic importance (Larkin et al., 2012).

The domestication of cattle offers a highly valuable model to study evolution; since their domestication, cattle have been continuously exposed to both natural and artificial selection, and these processes, coupled with genetic drift, have produced the multitude of breeds, phenotypes, and production characteristics evident today. Phenotypes selected by humans post-domestication include milk and meat production and quality, docility and temperament, fertility, and relatively aesthetic traits (such as coat color and polledness; Zeder et al., 2006). The selection processes influencing these phenotypes have left indelible marks on the genome of modern populations that can be detected via the analysis of data from high-throughput genotyping and sequencing platforms. For example, selection for beneficial genetic variants that control or contribute to a desired phenotype can result in the loss of sequence variation in the DNA flanking the selected variant and ultimately the fixation in the population of the chromosomal haplotype that harbors the selected variant. Chromosomal haplotypes that show fixation as a result of positive selection are said to be subject to a “selective sweep.” Analysis of selective sweeps can be used to identify regions of the genome that underlie important phenotypes. These traits have been subject to domestication-associated selective processes and, more recently, selection acting during breed formation and improvement, which may uncover the identity of genes and regulatory segments that have large phenotypic effects (Nielsen, 2005).

Evidence for selective sweeps has been detected in several pan-genomic surveys of bovine SNP diversity and resequencing projects. Results from the Bovine HapMap project identified selective sweeps centered around the myostatin (*MSTN*) and the ATP binding cassette sub family G member 2 (*ABCG2*) genes, which harbor SNPs associated with phenotypic differences in meat and milk production, respectively (Bovine HapMap Consortium, 2009). More recent studies have identified additional genomic regions subject to selective sweeps that encompass quantitative trait loci and candidate genes associated with beef and dairy production traits, behavioral traits, morphology, fertility, and health (Larkin et al., 2012; Qanbari et al., 2012; Ramey et al., 2013; Rothammer et al., 2013; Utsunomiya et al., 2013). It is important to note, however, that analysis of selective sweeps is challenging; for example, regions of the genome with significantly reduced sequence variation can also be generated by genetic drift, making true population genomic signals of selection difficult to discern fully.

Perhaps the most exciting application of high-throughput genomic technologies for investigating cattle domestication is through analysis of ancient pre-domestic whole genomes. In particular, NGS is revolutionizing paleogenetics, and considerable genomic DNA sequence information has already been reported from extinct mammals, including hominids (Miller et al., 2008, 2009; Green et al., 2010; Orlando et al., 2013). Until recently, reconstruction of a complete aurochs genome seemed unimaginable, but the first complete aurochs mtDNA genome sequence was published in 2010 (Edwards et al., 2010) using NGS technologies (in combination with conventional DNA sequencing). This sample was from an archaeologically verified British *B. p. primigenius* bone, radiocarbon-dated to 6,700 cal. YBP, more than 1,000 years before the arrival of domestic *B. taurus* to Britain (Rowley-Conwy, 2011). In addition to facilitating reconstruction of the mtDNA genome, this analysis also generated a large number of DNA sequences that mapped to the nuclear *B. taurus* genome, a finding that bodes well for generation of a complete *B. primigenius* genome sequence. Analysis of the genome sequence of the aurochs would provide a reference template for comparison of modern bovine genomes, providing the opportunity to more fully investigate the selective processes that have shaped modern cattle genomes since domestication more than 10,000 years ago. In this context, comparisons of both domestic and non-domestic genomes can answer fundamental questions about the evolutionary process and how it shapes biological variation in animal populations.

**Acknowledgements**

The authors wish to thank Dr. Johannes A. Lenstra for assistance with mtDNA sequence collation and analysis.


