Zebu Cattle Are an Exclusive Legacy of the South Asia Neolithic

Shanyuan Chen,†1 Bang-Zhong Lin,†2 Mumtaz Baig,3,4 Bikash Mitra,5,6 Ricardo J. Lopes,1 António M. Santos,1,7 David A. Magee,8 Marisa Azevedo,1 Pedro Tarroso,1 Shinji Sasazaki,2 Stephane Ostrowski,9 Osman Mahgoub,10 Tapas K. Chaudhuri,5 Ya-ping Zhang,6 Vânia Costa,1 Luis J. Royo,11 Félix Goyache,11 Gordon Luikart,1 Nicole Boivin,12 Dorian Q. Fuller,13 Hideyuki Mannen,2 Daniel G. Bradley,14 and Albano Beja-Pereira* 1

1Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, Vairão, Portugal
2Laboratory of Animal Breeding and Genetics, Graduate School of Agricultural Science, Kobe University, Kobe, Japan
3Department of Zoology, Government Vidarbha Institute of Science and Humanities, Amravati, Maharashtra, India
4State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China
5Cellular Immunology Laboratory, Department of Zoology, University of North Bengal, Siliguri, West Bengal, India
6Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, China
7Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal
8Animal Genomics Laboratory, UCD School of Agriculture, Food Science and Veterinary Medicine, UCD College of Life Sciences, University College Dublin, Belfield, Dublin, Ireland
9Wildlife Conservation Society, Bronx, New York
10Department of Animal and Veterinary Sciences, College of Agricultural and Marine Science, Sultan Qaboos University, Al-Khod, Muscat, Oman
11Area de Genética y Reproducción Animal, Servicio Regional de Investigación y Desarrollo Agroalimentario, Somió, Camino de los Claveles, Gijón, Spain
12Research Laboratory for Archaeology and the History of Art, University of Oxford, Oxford, United Kingdom
13Institute of Archaeology, University College London, London, United Kingdom
14Smurfit Institute of Genetics, Trinity College, Dublin, Ireland
*Corresponding author: E-mail: albanobp@fc.up.pt.
†These authors contributed equally to this work.
Associate editor: Connie Mulligan

Abstract

Animal domestication was a major step forward in human prehistory, contributing to the emergence of more complex societies. At the time of the Neolithic transition, zebu cattle (Bos indicus) were probably the most abundant and important domestic livestock species in Southern Asia. Although archaeological evidence points toward the domestication of zebu cattle within the Indian subcontinent, the exact geographic origins and phylogenetic history of zebu cattle remains uncertain. Here, we report evidence from 844 zebu mitochondrial DNA (mtDNA) sequences surveyed from 19 Asiatic countries comprising 8 regional groups, which identify 2 distinct mitochondrial haplogroups, termed I1 and I2. The marked increase in nucleotide diversity (P < 0.001) for both the I1 and I2 haplogroups within the northern part of the Indian subcontinent is consistent with an origin for all domestic zebu in this area. For haplogroup I1, genetic diversity was highest within the Indus Valley among the three hypothesized domestication centers (Indus Valley, Ganges, and South India). These data support the Indus Valley as the most likely center of origin for the I1 haplogroup and a primary center of zebu domestication. However, for the I2 haplogroup, a complex pattern of diversity is detected, preventing the unambiguous pinpointing of the exact place of origin for this zebu maternal lineage. Our findings are discussed with respect to the archaeological record for zebu domestication within the Indian subcontinent.

Key words: Bos indicus, domestication, pastoralism, neolithic, evolution, archaeology, anthropology.

Introduction

Plant and animal domestication, as part of new human productive strategies, represent arguably the most important global transformation in prehistory (Diamond 2005). The degree to which domesticates either spread to new areas from primary centers of origin or were independently developed at secondary locales during the Neolithic period is one of the major topics of archaeological and, more
recently, genetic investigations. Current evidence suggests that the Indian subcontinent witnessed both the dispersal of domesticates from agricultural centers situated further west (such as the Fertile Crescent) and the indigenous domestication of local species (Fuller 2006). One of the key Neolithic centers of the Indian subcontinent was undoubtedly the Baluchistan region (situates in present-day Pakistan), where the arrival of new crops from the Near East ~9,000 years before present (YBP) are thought to have prompted the domestication of more localized wild progenitor species, including the South Asian aurochs, Bos primigenius namadicus—the purported ancestor of modern zebu cattle (Bos indicus) (Grigson 1980; Jarrige and Meadow 1996). It was supposed that B. primigenius namadicus ranged over the Indian subcontinent during Pleistocene and Holocene periods and that some of their populations almost certainly survived into Neolithic times to give rise to B. indicus (Grigson 1985; Van Vuure 2005). Evidence retrieved from the archaeological sites of Harappa and Mohenjo-daro indicates that domestic zebu were widespread throughout the Indus Valley region ~5,000 YBP (Meadow 1993, 1996; Fuller 2006).

More recently, South India has also been proposed as another independent center of domestication within South Asia, specifically for crops (Fuller 2006). Moreover, the observed morphological differences between cattle depicted in the rock art of South India and in the iconography of Indus Valley civilizations have also led to suggestions that South India was a secondary center for zebu domestication (Allchin FR and Allchin B 1973). This hypothesis is supported by the presence of a distinctive, cattle-oriented Neolithic culture in South India that produced hundreds of unique ashmounds (mounds of burnt cattle dung), but archaeological data confirming such a scenario are lacking. Recent zooarchaeological data suggest that wild cattle were present in the region into the Neolithic period (Korisetar et al. 2001). Other potential centers of zebu domestication, which likely featured a combination of allochthonous and autochthonous processes, include Gujarat and the Ganges region, where, according to archaeological data, domestic zebu were present ~5,500 and ~4,000 YBP, respectively (Fuller 2006). Small numbers of bones from both regions suggest the persistence of wild cattle. Today, the majority of domestic cattle from Europe and North Eurasia are humpless taurine-like (Bos taurus), whereas humped zebu cattle predominate in South Asia and Southeast Asia. Zebu cattle are also encountered in South China where they are believed to have been introduced from domestication centers situated further west some 2,500 YBP (Higham 1996). In contrast, taurine cattle are believed to have spread from Central Asia to Central and Northern China between 5,000 and 4,000 YBP (Flad et al. 2007).

Although investigations of mitochondrial DNA (mtDNA) sequence variation have confirmed the independent domestic origins of B. taurus and B. indicus cattle from genetically divergent wild aurochs progenitors (Loftus et al. 1994) and have shed much light on the ancestry of B. taurus cattle (Troy et al. 2001; Beja-Pereira et al. 2006), studies involving zebu mtDNA are limited. Previous phylogenetic studies have shown that zebu mtDNA sequences cluster into two distinct groups each consisting of a centrally positioned, numerically predominant (and hence presumably ancestral) sequence (termed the I1 and I2 haplotypes), through which all derivative haplotypes coalesce (Baig et al. 2005; Lai et al. 2006; Magee et al. 2007). The star-like patterns of diversity within these sequence groups (herein referred to as the I1 and I2 haplogroups) are analogous to the patterns of diversity revealed for B. taurus mtDNA sequences (Troy et al. 2001) and are indicative of historic population expansions, presumably associated with the domestication process itself. However, due to restricted sampling within South and East Asia, the patterns of mtDNA sequence diversity and geographical partitioning of the I1 and I2 haplogroups have not been fully resolved (Baig et al. 2005; Lai et al. 2006; Magee et al. 2007).

To investigate whether zebu cattle were domesticated once or several times, and whether such domestication occurred exclusively within the Indian subcontinent, we analyzed 844 zebu mitochondrial control region sequences surveyed from 19 countries distributed throughout West Asia, South Asia, and East Asia comprising 30 discrete populations which were further grouped into eight major geographic regional groups (Supplementary table S1, Supplementary Material online). Phylogenetic analysis of the data reveals 94 distinct mtDNA haplotypes differentiated at 60 polymorphic sites. The predominant haplotype (I1) was observed 391 times, whereas the second most frequent haplotype (I2) was observed 118 times. All remaining haplotypes fall into the two previous defined I1 and I2 haplogroups, with a mean internal divergence of 3.42 nucleotides (i.e., corrected mean pairwise differences). Of the total 94 detected haplotypes, 56 fall within the I1 haplogroup (607 sequences) and 38 are encountered within the I2 haplogroup (237 sequences). Overall, the mean number of pairwise differences within the I2 haplogroup (1.411 ± 0.867) is higher than that for the I1 haplogroup (1.143 ± 0.742).

To assess the partitioning of zebu mtDNA diversity across Asia, we separately analyzed nucleotide diversity (π; Nei 1987) levels within the I1 and I2 haplogroups for all eight defined regional groups, including Indus Valley (Pakistan and India), Ganges (India and Bangladesh), South India, Northeast Indian subcontinent (Bhutan, Nepal and India), West Asia (Iraq, Oman, and Turkey), Central Asia (Kazakhstan, Kyrgyzstan, Turkmenistan, and Afghanistan), East Asia (China and Mongolia), and Southeast Asia (Myanmar, Cambodia, Laos, Vietnam, and Philippines) (Supplementary table S1, Supplementary Material online). Notably, increased nucleotide diversity was observed for all four geographic regions within the Indian subcontinent, namely the 1) Indus Valley region, 2) the Ganges region floodplains bordered by the Brahmaputra River, 3) South India, and 4) Northeast Indian subcontinent—compared with all other geographic regions (P < 0.001) (fig. 1 and table 1). Modern populations from centers of domestication are expected to display elevated levels of genetic diversity due to an increased retention of captured, wild
genetic variation. Furthermore, as the formation of populations outside these centers would have undoubtedly involved the subsampling of this ancestral variation, genetic diversity in modern populations generally decreases with increasing distance from the center of origin—an observation documented in previous studies (Troy et al. 2001; Beja-Pereira et al. 2004). Hence, the nucleotide diversity estimates for the \( I_1 \) and \( I_2 \) haplogroups presented here are consistent with the Indian subcontinent having served as the center of origin for modern domestic zebu cattle. Furthermore, we used bootstrap tests of significance (Manly 1997) to test for the effect of regional sample size on genetic diversity. For each group, we generated 10,000 replicates using sample sizes of \( n = 40, 100, \) and \( 200 \). The \( P \) values were estimated as the fraction of bootstrap samples in which nucleotide diversity was lower than the observed one. In all cases, the observed genetic diversity fell within the 95% interval of the bootstrap distribution \( (P > 0.05) \), suggesting that differences in sample size do not affect the estimates of genetic diversity presented here.

To further determine the location of the probable center for domestication of the two zebu mtDNA haplogroups, among the three main purported regions of zebu domestication, namely the Indus Valley, the Ganges region, and South India (Allchin 1963; Fuller 2008), we performed detailed comparative analyses of diversity for each haplogroup. These analyses allowed us to identify the greater Indus Valley (including Rajasthan and present-day Pakistan) as the most likely location for the origin of the \( I_1 \) haplogroup and, hence, a strong candidate for

---

**Fig. 1.** Geographic distribution compilation between mtDNA genetic patterns across Asia and the archaeological signs of spread of cattle pastoralism within the Indian subcontinent. (A) Median-reduced networks constructed for zebu haplotypes across Asia; (B) a map of the Indian subcontinent showing median reduced networks for each potential domestication center (Indus, Ganges, and South India); (C) a map of the Indian subcontinent indicating the spread of cattle across time based on archaeological data. Circles represent sites containing domesticated zebu cattle faunal remains, and squares represent reports of Holocene wild-type cattle bones. For dates, see Supplementary table S3 (Supplementary Material online).
the primary center of zebu domestication, as indicated by compelling archaeological evidence (Meadow 1984, 1996). Indeed, this haplogroup is by far the most frequent, not only in India but also the rest of the Asia (fig. 1) and, in some regions, is the only observed mitochondrial lineage. The Indus Valley as a primary center of origin of zebu domestication is supported by three lines of genetic evidence: 1) the populations of this Indus Valley region display a complete backbone structure for the network of the I1 haplogroup (fig. 1 and Supplementary fig. S1, Supplementary Material online); 2) the modal value of the mtDNA mismatch distribution for the Indus Valley region (modal value 2) is higher than the modal value generated for the I1 mismatch distributions of the Ganges and South Indian regions (modal value 1), suggesting that expansion of the I1 haplogroup in the Indus region predates that within the Ganges and South India regions (fig. 2); and 3) the Indus region displays a larger number of unique haplotypes for this haplogroup compared with both the Ganges and South Indian regions (Supplementary table S2, Supplementary Material online).

However, when similar analyses of the data were performed for the I2 haplogroup, a more complex pattern of diversity was detected, whereby some haplotypes derived from the presumed ancestral I2 haplotype were observed at high frequencies, whereas other derivative haplotypes were encountered only in specific locales. The most notable case is an intermediate I1–I2 haplotype, designated I2a (and considered here as a member of the I2 haplogroup), which is differentiated from the I1 and I2 central haplotypes by three- and one-nucleotide substitutions, respectively (fig. 1). For example, this haplotype displays a high frequency (93%) in the Uttar Pradesh region (Supplementary fig. S1, Supplementary Material online), in the middle plains of the Ganges, with a corresponding reduction in frequency away from this region. This striking pattern may indicate that during expansion of cattle pastoralism, local wild females were recruited into domestic populations in Ganges Valley and central India, in a diffused geographical and temporal process.

Furthermore, almost equal degrees of increased diversity, illustrated by similar network structures (fig. 1 and Supplementary fig. S1, Supplementary Material online), identical modal values of mtDNA mismatch (modal value 1, fig. 2), and nearly identical numbers of unique haplotypes (Supplementary table S2, Supplementary Material online) for the I2 haplogroup, were found in both the Indus Valley and South Indian regions. These results, together with the high frequency of the unique I2a haplotype found in the Ganges (Supplementary fig. S1, Supplementary Material online), prevent the unambiguous pinpointing of a single region for the origin of the I2 haplogroup. Archaeological

**Table 1.** Test for Differences of Nucleotide Diversity (π for I1 haplogroup, π for I2 haplogroup) between the Indian Subcontinent and All Other Locations.

<table>
<thead>
<tr>
<th>Haplogroup</th>
<th>Location</th>
<th>n</th>
<th>Mean π</th>
<th>Standard deviation</th>
<th>U</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>Indian subcontinent</td>
<td>14</td>
<td>0.00700</td>
<td>0.00200</td>
<td>224 &lt;0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>I2</td>
<td>Indian subcontinent</td>
<td>14</td>
<td>0.00530</td>
<td>0.00290</td>
<td>223 &lt;0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Other</td>
<td>Indian subcontinent</td>
<td>16</td>
<td>0.00210</td>
<td>0.00130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Indian subcontinent</td>
<td>16</td>
<td>0.00010</td>
<td>0.00040</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.**—Other geographic regions include West Asia, Central Asia, East Asia, and Southeast Asia. “n” is the number of populations. Hypothesis tested using nonparametric tests (Mann–Whitney [U] test). Significance levels are symbolized by (∗P).

**Fig. 2.** Mismatch distributions and tests of neutrality for the three hypothesized centers of zebu origin (Ganges, Indus, and South India).
evidence for domesticated zebu is earlier in the Indus Valley (~8,000 YBP) than South India (~5,000 YBP) and middle Ganges (~4,000 YBP), and parallels geographically the distribution of probable wild aurochs (see fig. 1C, Chattopadhyaya 2002; Fuller 2006), suggesting that the I2 haplogroup was initially more likely domesticated in the northern part of the Indian subcontinent rather than South India and middle Ganges. However, commercial exchanges between the societies of the Indus Valley region and those of South India and the Ganges region from the Neolithic to the present day may have erased any ancient phylogeographic DNA signature of haplogroup domestications.

The dramatic predominance of the I1 haplogroup across Southeast Asia and high frequency in China is also noteworthy and may suggest that this lineage arrived with the first domestic herds from India. Archaeological evidence suggests that cattle arrived in South China and Southeast Asia no earlier than 2,000 BC and, perhaps, closer to 1,500–1,000 BC (Higham 1996). The near absence of I2 in Southeast Asian zebu would seem to suggest a later incorporation of the I2 haplogroup into the domestic pool, perhaps, during the Chalcolithic or Iron Age period. Culturally established breeds in China and Southeast Asia by this time would have precluded the simple spread of I2, whose scattered presence in the region is, perhaps, more readily explained as resulting from later diffusion by trade.

Applying the mutation rate of 38% per million years per site by Troy et al. (2001), we can obtain the time since the expansion for each haplogroup. The expansion times, expressed in mutational units, t, were estimated to be 2.496 (confidence interval [CI] 90% = 0.000–4.438) and 1.496 (CI 90% = 0.346–2.865), for I1 and I2 haplogroups, respectively. The approximate time since the expansion was therefore 13,600 years for I1 haplogroup and 8,200 years for I2 haplogroup. The expansion times of I1 and I2 haplogroups are, therefore, concordant with our hypothesis of two expansion processes for zebu domestication.

In summary, our data are indicative of the domestication of zebu cattle exclusively within the northern part of the Indian subcontinent. Furthermore, although our genetic data corroborate archaeological inferences that the Indus Valley was most likely the primary center of zebu domestication, the frequency and distribution of the I2 haplogroup within Uttar Pradesh and the Ganges region tentatively suggest, at least, a secondary recruitment center of wild female aurochs into proto-domestic zebu within Northern India. Under this scenario, members of the I1 haplogroup (perhaps occasionally the I2 haplogroup) were first adopted into the domestic pool during the early phases of zebu domestication in the Indus Valley (~8,000 YBP), which was undoubtedly pivotal to the emergence of pastoralism throughout India (~5,500–4,000 YBP) and its diffusion eastward toward Southeast Asia and Southern China (<4,000 YBP). Sometime after this initial spread, additional genetic diversity belonging to the I2 haplogroup was recruited from local wild South Asian populations, perhaps, as wild cattle populations were going extinct. This was, perhaps, the most intensive in the Ganges region, compatible with archaeological evidence for the presence of wild aurochs in the Ganges in Neolithic times (see fig. 1C, Chattopadhyaya 2002; Fuller 2006) but may also have occurred on a lesser scale in Southern India. Although available evidence for the late distribution of wild aurochs in South India is scarce, the possibility remains that some wild aurochs may have survived into Neolithic times in South India, given the proposed continent-wide distribution of wild aurochs during the late Pleistocene and Holocene (Grigson 1985; Van Vuure 2005). Remains of wild aurochs (B. primigenius namadicus), dated 2,200 BC, have been clearly identified from Banahalli, Kanataka (see fig. 1C and Supplementary table S3, Supplementary Material online) (Korisettar et al. 2001; Chattopadhyaya 2002; Fuller 2006). Parallel recruitment of wild bovines may be postulated for Southeast Asia in the domestication of other sister species, such as Bos gaurus and Bos javanicus, perhaps, by ~3,500 YBP, but their domestication is poorly documented.

We have demonstrated, once more, that mtDNA sequencing analysis studies not only complement archaeological evidence but also add information by linking some divergent lineages/haplogroups to geographic origins and directions of spread. Identification of dispersal history and centers of origin may help reveal potential sources of genetic diversity to be conserved and used for future improvement of livestock and agricultural production. This is important as future productivity and adaptation to environmental changes might be overcome using zebu crosses, as pastoralists did in ancient times.

Materials and Methods

Sampling, DNA Extraction and Sequencing and GenBank Sequence Mining

Tissue samples of 548 local zebu cattle individuals were collected from 16 countries across Asia (Supplementary tables S4, Supplementary Material online). We only sampled small countryside villages and excluded research centers, large cities, and coastal harbors where recent shipping of cattle might be possible. Efforts were made to avoid sampling related individuals. Genomic DNA was extracted by DNeasy Blood & Tissue Kit (QIAGEN GmbH, Hilden, Germany). A 320-bp fragment of mtDNA control region hypervariable region 1 was amplified and sequenced by using primers AN2_for and AN3_rev (Troy et al. 2001), following procedures described before (Beja-Pereira et al. 2006). Raw sequences were checked and aligned using DNASTAR 6.0 (DNASTAR, Inc., Madison, WI). The resulting sequences were aligned with 296 sequences of Asiatic zebu cattle from GenBank database (Supplementary table S4, Supplementary Material online) and finally generated a data set of 844 sequences in length of 240 bp, between 16,023 and 16,262 of reference mtDNA genome sequence V00654. This final data set covers 19 Asiatic countries.

Data Statistical Analysis

Mean number of pairwise differences, and nucleotide diversity (π), neutrality tests such as Tajima’s D and Fu’s Fs, mismatch distribution, and raggedness index were calculated
by Arlequin 3.1 (Excoffier et al. 2005). Network profiles among haplotypes were constructed by median-joining networks (Bandelt et al. 1999) (NETWORK 4.5; http://www.fluxus-engineering.com/), resolving the reticulations through a maximum parsimony criterion. The time of the expansion for each haplogroup was estimated by using the parameters of the demographic expansion (i.e., mismatch distribution), \( \tau \), expressed in mutational units (\( \tau = 2ut \), where \( u \) is the mutation rate for the whole sequence and \( t \) is the time since the expansion). The \( \tau \) values and their CIs were obtained by 10,000 bootstrap replications in Arlequin 3.1 (Excoffier et al. 2005). The bootstrap tests of significance to test for the effect of sample size on genetic diversity was constructed following the method suggested by Manly (1997).

### Supplementary Material

Supplementary figure S1 and supplementary tables S1–S4 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/). All new sequences produced in this study have been deposited in GenBank under accession numbers FJ492194–FJ492741.

### Acknowledgments

The authors thank two anonymous reviewers for their helpful suggestions and comments and also thank Ceiridwen Edwards for the comments on an early version of the manuscript. This work was funded by Fundação para a Ciência e Tecnologia (FCT) project POCTI/CTV/56758/2004. S.C., R.J.L., and P.T. are supported by FCT individual grants SFRH/BD/42480/2007, and SFRH/BD/42480/2007, respectively. G.L. was supported by the Portuguese-American Foundation for Development, Centro de Investigação em Biodiversidade e Recursos Genéticos, and University of Porto. F.G. and L.J.R. were funded by the grant CGL2005-03761/BOS. Y.P.Z. was supported by National Natural Science Foundation of China and Yunnan Province, China. Collection of samples in Afghanistan has been done with the financial support of United States Agency for International Development. We also thank Tamal Mazumder, Debojit Saha, Chhotan Ghosh, and M. D. Eusuph for their support and help with sampling in India.

### References


