Phylogeography and origin of Indian domestic cattle

Cattle had a pivotal role in the up-liftment of human culture and civilization. Besides being used mainly as a beast of burden and ceremonial animal in prehistoric and historic time, cattle were also used as a rich source of protein and fat. The present-day domestic cattle are broadly classified into two groups: zebu (humped) and taurine (humpless), which are scientifically referred as Bos indicus and Bos taurus respectively. A clear dichotomy has been described for zebu and taurine mitochondrial DNA (mtDNA). This dichotomy can only be explained by the presence of two subspecies of wild aurochs before domestication. The Indian subcontinent harbours a variety of cattle. There are evidences that the ~5000 yr old ancient Indus civilization had made marked progress in the field of animal husbandry. Besides many non-descript breeds, there are 30 well-recognized cattle breeds in India. Majority of these breeds are low producers of milk; hence they are primarily used for the production of castrated bulls, which are used in agriculture, carting and transport. Previous studies on Indian cattle have proved independent domestication centre for Indian zebu. No attempt has been made to measure the time depth for this important event. In view of the above data, we attempted to undertake a pilot study using long-range mtDNA marker to obtain vital insights into the domestication of zebu cattle. Fresh blood samples from 25 local cattle were collected from three geographical regions (Figure 1). The local cattle fall under the non-descript category is considered as the founder population for the breed. According to the information provided by farmers, precaution was taken to collect samples from unrelated individuals from countryside villages. DNA was isolated from leucocytes by standard phenol chloroform extraction methods.

Using polymerase chain reaction, partial mtDNA D-loop regions were amplified. Primers were designed to amplify a 375-bp fragment from the hypervariable region 1 (HVR-1), AN4-bio (L15960; 5′-GGTA-ATGTACATAACATTAATG-3′) and AN3 (H16334; 5′-CGAGATGTCTTATTTA-AAGGG-3′). Primer AN4-bio was biotinylated for subsequent product purification. Reactions were performed using 10 ng of template mtDNA in a 50 μl reaction volume, 2.5 units of Taq DNA polymerase, 10X reaction buffer (50 mM KCl; 10 mM Tris-HCl pH 9.0; 1.0 Triton X-100, 1.5 mM MgCl2), with concentrations of 200 mM for each dNTP and 2 ng μl-1 of both the forward and reverse primers. Before amplification a 20 μl oil overlay was added to each sample. Amplification was done using a 4 min denaturation step followed by 40 cycles of 40 s at 94°C, 40 s at 55°C, 40 s at 72°C and a final extension at 72°C for 4 min. Reaction products were purified using Dynabeads (DYNAL), according to the manufacturer’s instructions. Sequencing of a 240-bp fragment (16023–16262) for both strands was performed using the dideoxy chain termination method. The sequences were deposited in GenBank (accession no. AJY972130–AJY972154).

Using Clustal X, 240-bp sequences were aligned with Anderson reference sequence. These 25 sequences were analysed in conjunction with previously published sequences of cattle from North and South India. In order to broaden the study, published sequences were purified using Dynabeads (DYNAL), according to the manufacturer’s instructions. Sequencing of a 240-bp fragment (16023–16262) for both strands was performed using the dideoxy chain termination method. The sequences were deposited in GenBank (accession no. AJY972130–AJY972154).

Using Clustal X, 240-bp sequences were aligned with Anderson reference sequence. These 25 sequences were analysed in conjunction with previously published sequences of cattle from North and South India. In order to broaden the study, published sequences were purified using Dynabeads (DYNAL), according to the manufacturer’s instructions. Sequencing of a 240-bp fragment (16023–16262) for both strands was performed using the dideoxy chain termination method. The sequences were deposited in GenBank (accession no. AJY972130–AJY972154).

Using Clustal X, 240-bp sequences were aligned with Anderson reference sequence. These 25 sequences were analysed in conjunction with previously published sequences of cattle from North and South India. In order to broaden the study, published sequences were purified using Dynabeads (DYNAL), according to the manufacturer’s instructions. Sequencing of a 240-bp fragment (16023–16262) for both strands was performed using the dideoxy chain termination method. The sequences were deposited in GenBank (accession no. AJY972130–AJY972154).
of taurine cattle from Near East, Europe and Africa were also included. The nucleotide and gene diversity was computed using DnaSP 4.0 version. Mismatch distribution and median joining network were constructed using NETWORK 4.1 version. The coalescence age was estimated using the estimator $\rho$, which is the average transitional distance from the most recent common ancestor.

From the estimated diversity indices, it becomes evident that cattle from Near East exhibit highest diversity (Table 1). Similarly, taurine cattle exhibit high haplotype diversity compared to zebu cattle. In terms of nucleotide diversity, except Africa and Europe, all groups show high value. The overall nucleotide diversity was found to be nearly same for both zebu and taurine cattle (Table 1).

Keeping in order with the previously studied taurine population, the network constructed using median-joining algorithm exhibits two star-like expansion events radiating from two ancestral nodes designated as Z1 and Z2 (Figure 2a and b). The approach reveals distinct dichotomy between zebu and taurine cattle encompassing large separation time (Figure 2a). This demography is further explored in subsequent mismatch distribution and coalescence age estimation.

We subjected the sequence information to a robust statistics developed over a year, known as pairwise differences (also referred to as mismatch distribution analysis). Interestingly, two smooth, bimodal distributions separated by a large time interval are obtained (Figure 3). Further, the Fu’s $F_s$ statistics, which is sensitive to population growth, yields significant departure from neutrality in each group. In a marked display of symmetry with the Near East, African and European samples, the zebu cattle too show star-like expansion. Within the 95% CI, estimation of time depth to Z1 and Z2 clusters using $\rho$ statistics and mutational rate (38% per m.y.), yields 9978 and 12418.

### Table 1. Estimated diversity and demographic indices

<table>
<thead>
<tr>
<th>Group</th>
<th>No. of sequences</th>
<th>Gene diversity</th>
<th>Nucleotide diversity</th>
<th>Fu’s $F_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>South India</td>
<td>21</td>
<td>0.92</td>
<td>0.017</td>
<td>-4.80</td>
</tr>
<tr>
<td>North India</td>
<td>23</td>
<td>0.96</td>
<td>0.016</td>
<td>-10.53</td>
</tr>
<tr>
<td>Africa</td>
<td>25</td>
<td>0.90</td>
<td>0.009</td>
<td>-17.64</td>
</tr>
<tr>
<td>Europe</td>
<td>25</td>
<td>0.95</td>
<td>0.009</td>
<td>-22.48</td>
</tr>
<tr>
<td>Near East</td>
<td>25</td>
<td>0.98</td>
<td>0.019</td>
<td>-15.57</td>
</tr>
<tr>
<td>Taurine</td>
<td>75</td>
<td>0.97</td>
<td>0.017</td>
<td>-76.46</td>
</tr>
<tr>
<td>Zebu</td>
<td>50</td>
<td>0.93</td>
<td>0.017</td>
<td>-24.48</td>
</tr>
</tbody>
</table>

![Figure 2.](image1.png) **Figure 2.** a. Median-joining network constructed from *Bos indicus* and *Bos taurus* cattle sequences (size of node is proportional to haplotype frequency). Enlarged median-joining network, exhibiting (b) *B. indicus* and (c) *B. taurus* clusters.

![Figure 3.](image2.png) **Figure 3.** Mismatch distribution constructed using *B. indicus* and *B. taurus* sequences.
years for Z1 and Z2 clusters respectively (Figure 2a). Moreover, the three previously studied taurine clusters in the lower branch give time depths of 11768, 10928, and 8904 yrs respectively, for Europe, Near East and Africa (Figure 2c).

The high and low gene diversity for taurine and zebu cattle respectively, was corroborated by their geographical distribution. Taurine cattle are more widespread in their distribution when compared to zebu. The occurrence of high nucleotide diversity in zebu cattle, similar to those from Near East, could be due to separate domestication for *B. indicus* from separate subspecies of wild auroch (Table 1). The median joining network also shows two star-like expansions for zebu, suggestive of population expansion19. Notably, in the phylogenetic network, the zebu cattle were characterized by two of the oldest nodes, designated as Z1 and Z2. Both these nodes harbour cattle from South India (Figure 2a). The occurrence of such cattle sequences in both the ancestral nodes indicates that they might have provided the inocula for early domestication events probably starting from South India. In Figure 3, the two well-separated bimodal mismatch distributions for *B. indicus* and *B. taurus* can only be explained by independent domestication events for both the cattle breeds. This demographic history was further supported by estimates of Fu’s Fs values, which are sensitive to population expansion (Table 1). The time-depth estimates around the two zebu clusters are in agreement with the history of cattle domestication (Z1 and Z2 yield 9978 and 12418 yrs respectively). The Fs values and coalescence age estimates indicate that the origin and spread of agropastoristolic societies would have resulted in drastic and sustained population increase in both the early herders and their flock. These patterns are probably indicative of genetic signature of expansion after domestication. Further, Dravidian-speaking people dominate South India. There are conflicting views regarding the origin of Dravidian languages in India. According to some scholars, the Dravidian language originated in India18,20, while others proposed that the Elamo-Dravidian languages originated in the Elam province of southwestern Iran, and the dispersal of the Dravidian languages into India took place with human migration from this region21,22. Accordingly, the ancestors of present-day Dravidians brought with them the technologies of agriculture and animal domestication, thereby supporting demic expansion23. Our estimates of time to the most recent common ancestor for two zebu clusters clearly indicate Neolithic transition. These are potentially the genetic signals of independent cattle domestication in India, in parallel to earlier suggested Near Eastern domestication for European and African cattle24. In future, probing by Y-chromosome STR marker will shed more light on the unequal gene flow associated with such kind of single locus study.


Received 4 January 2005; revised accepted 12 April 2005

**MUMTAZ BAIG**1,a
**ALBANO BEJA-PEREIRA**2
**RIZWAN MOHAMMAD**3
**KASHINATH KULKARNI**1
**SAMIRA FARAH**1
**GORDON LUIKART**2

1Population Biology Laboratory, Department of Zoology, Government Vidarbha Institute of Science and Humanities, Amravati, India
2Lab. d’Ecologie Alpine, UMR CNRS-UJF 5533, 38041 Grenoble, France
3Department of Biochemistry, Shri Shivaji Science College, Akola, India
*aFor correspondence. e-mail: samitramumtaaz@yahoo.com**