

# Admixture analysis of South Asian cattle

P Kumar<sup>1,5</sup>, AR Freeman<sup>2,5</sup>, RT Loftus<sup>2</sup>, C Gaillard<sup>3</sup>, DQ Fuller<sup>4</sup> and DG Bradley<sup>2</sup>

<sup>1</sup>Animal Genetics Department, Indian Veterinary Research Institute, Izatnagar, Bareilly, UP, India; <sup>2</sup>Department of Genetics, Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland; <sup>3</sup>Institute of Animal Breeding, University of Bern, Bern, Switzerland; <sup>4</sup>Institute of Archaeology, University College London, 31-34 Gordon Square, London WC1H 0PY, UK

We present population genetic analysis of microsatellite variation in seven *Bos indicus* cattle breeds from a variety of locations in South Asia. This is the first such study focusing within this area, which is one of the postulated centres of origin of domestic cattle. An estimate of the influence of *Bos taurus* ancestry was carried out using three approaches: by the systematic selection of population-associated alleles for *B. taurus* and examination of their frequency; by examining

the truncation of genetic distances from European populations; and by a model-based Bayesian admixture analysis. These analyses revealed a *B. taurus* influence in the Indian subcontinent; part of a gradation which stretches from Europe through the Near East towards Indian and which may be of ancient origin.

*Heredity* (2003) 91, 43–50. doi:10.1038/sj.hdy.6800277

**Keywords:** domestication; admixture; *Bos indicus*; microsatellite; stepwise

## Introduction

Genetic and archaeological evidence supports at least two domestications of cattle from different wild progenitor aurochs races. *Bos indicus*, or zebu, are postulated to have domestic origins in the Baluchistan region, while *Bos taurus*, also termed taurine cattle, arose in the Near East and possibly North Africa. Genetic studies have estimated the divergence time between ancestral strains at several hundred thousand years (Loftus *et al.*, 1994; MacHugh *et al.*, 1997).

To re-examine the origins of zebu cattle, we here analyse microsatellite allele frequency data from 20 markers that have been screened in seven breeds with origins in a variety of locations in South Asia. These data represent the first molecular survey of cattle within the region and support earlier findings which suggest that zebu were domesticated from a separate strain of aurochs to that which gave rise to *B. taurus* (taurine) cattle (Loftus *et al.*, 1994). The nature of any secondary contact between zebu cattle and other domestic strains is of interest from a historical perspective. It also has important implications for any admixture studies, where Indian cattle are often used as representative of putative parental populations (Loftus *et al.*, 1999; Hanotte *et al.*, 2002).

## Admixed populations

Several studies have been carried out to examine the amount and pattern of admixture in selected human populations (Parra *et al.*, 1998; Wilson and Goldstein, 2000; Pfaff *et al.*, 2001; Collins-Schramm *et al.*, 2002). A serious obstacle to these studies is the genetic similarity

of all human groups and the potential difficulty in finding private or population-specific alleles (McKeigue, 1997; Shriver *et al.*, 1997). When present, these alleles may be assigned to an individual parental population and this enables the quantification of the amount of introgression from a parental population in a hybrid sample. Moreover, if alleles are neither fixed in the parental population nor absent in the comparison populations, their frequency differential may still enable their use to infer population history. They are thus referred to as ancestry-informative or population-associated alleles (PAAs).

The *B. indicus* and *B. taurus* genomes are sufficiently different such that there are many both population-specific and population-associated microsatellite alleles. The former have been used, after identification by eye, to quantify the amount of zebu introgression into native taurine populations in Africa and the Near East. Mitochondrial DNA and Y-chromosome and other allele frequency-based approaches have been used to support admixture estimation by zebu-specific alleles. In these studies, European and African taurine and Indian zebu cattle have been used as parental populations with which to infer the amount of admixture in hybrid breeds (Bradley *et al.*, 1994; MacHugh *et al.*, 1997; Loftus *et al.*, 1999; Hanotte *et al.*, 2000). However, the accuracy of these methods depends on the availability and correct selection of parental populations. There is a reasonable hypothesis, given its proximity to the Eastern region of the Fertile Crescent, that *B. taurus* admixture (possibly very early in origin) may have encroached into India. Without a description of this, there is no objective method with which to choose a parental (or introgressed) *B. indicus* population without using circular argument.

Here, we supplement new microsatellite data with published data and have devised an unbiased system to select taurine-specific alleles. We have examined the frequencies of these alleles in seven Indian breeds to ascertain the degree, if any, of taurine introgression. We

Correspondence: DG Bradley, Department of Genetics, Smurfit Institute of Genetics, Trinity College, Dublin 2, Ireland. E-mail: dbradley@tcd.ie

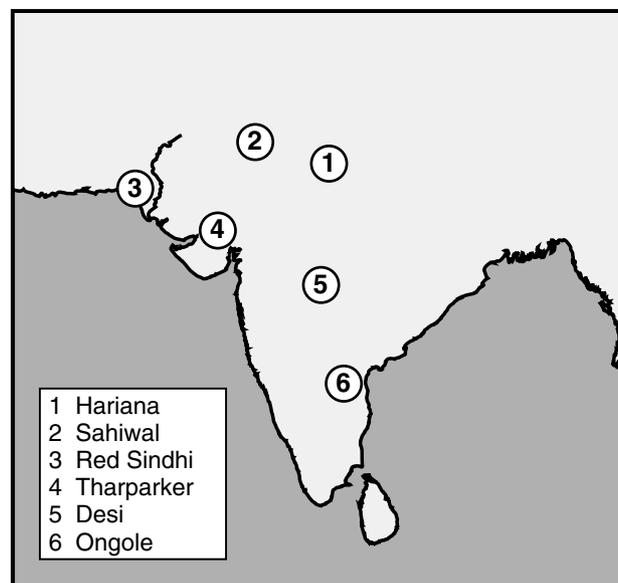
<sup>5</sup>Both equally contributed to this work.

Received 29 May 2002; accepted 18 February 2003

have tested the authenticity of our allele choice and admixture estimates. Additionally, we examined the average genetic distance between each Indian breed and 11 European populations as an alternative indication of *B. taurus* influence. Finally, Bayesian methods developed by Pritchard *et al* (2000) were employed to estimate the posterior distributions of individual admixture. This technique, which does not require parental population identification, has been employed in similar studies in wildcats (Beaumont *et al*, 2001) and brown trout (Hansen *et al*, 2001), and here illustrates the taurine influence in Indian *B. indicus* cattle. We have avoided the use of traditional allele-frequency-based admixture measures that require preselection of parental populations, and the inherent bias that they bring to admixture analysis.

## Materials and methods

To augment samples collected for other studies (Loftus *et al*, 1994; Edwards *et al*, 2000; Ritz *et al*, 2000), hair samples were collected from cattle breeds in the South Asia region (Table 1; Figure 1) ensuring, when possible, that the animals selected were unrelated and considered typical of the breed. There is a considerable debate about how to classify domestic descendants of a wild species. In this paper, we follow the Linnæan taxonomy with two domesticated species, the taurine cattle *B. taurus* and the zebu cattle *B. indicus*, while fully recognising that the two species are fully interfertile and therefore sometimes considered as subspecies by some authors (Hanotte *et al*, 2002). Desi is a Hindi name used for nondescript local cattle, but for the purposes of the following analyses these animals will be referred to as a breed. Sampling locations are shown in Figure 1. The Nellore breed was sampled in Brazil and is derived from Ongole cattle imported from India from 1895 onwards (Felius, 1995); data from both have been represented previously (Loftus *et al*, 1999). DNA from hair was extracted and samples were screened with 20 microsatellite loci using the procedures outlined elsewhere (Loftus *et al*, 1999). The loci have been selected from a list of 30 recommended for collaborative study ([http://www.ri.bbsrc.ac.uk/cdiv\\_www/homepage.htm](http://www.ri.bbsrc.ac.uk/cdiv_www/homepage.htm)) and were: CSRM60, CSSM66, ETH3, ETH10, ETH152, ETH225, HEL1, HEL5, HEL9, HEL13, BMS1818, BMS1824, BMS2113, ILST005, ILST006, INRA005, INRA023, INRA032, INRA035, and INRA063. The European (*B. taurus*) breeds that were used are: Jersey (Jer), Simmental (Sim), Holstein (Hol), Swiss Brown (OB), Eringer (Eri), Evo-



**Figure 1** Map of the Indo-Pakistan region showing the site of origin of six Indian cattle breeds. The Nellore breed was sampled in Brazil but is derived from the south Indian Ongole breed. Desi are local nondescript cattle that were sampled at this location and do not represent an official breed.

lander (Evo), Charolais (Cha), Vosges (Vog), Pustertaler (PS), Pinzgauer (PG), and Hungarian Grey (HG). The Near Eastern breeds that were used are: Turkish Grey (TG), Anatolian Black (ABl), South Anatolian Red (SAR), East Anatolian Red (EAR), Damascus (Dam), Kurdi (Kur), and Iraqi (Irq).

Allele frequencies for all loci were determined directly by counting. Unbiased estimates of gene diversity (expected heterozygosity) with associated standard errors were computed for all locus/population combinations according to Nei (1987). The mean number of alleles per locus (MNA) was also calculated as an index of sample diversity. A size-adjusted MNA value was calculated using 1000 iterations of a resampling method.

*B. taurus*-associated alleles were selected by comparing pooled European breed data with pooled data from all Indian samples. A scaled frequency differential (SFD) was calculated for each allele by the formula

$$\frac{|P_E - P_I|}{\max(P_E, P_I)}$$

**Table 1** Indian breed names, sample sizes, heterozygosity, and MNA measures

| Breed name | Code | n  | Expected heterozygosity | Exp Hz SD | MNA  | Sample size-adjusted MNA |
|------------|------|----|-------------------------|-----------|------|--------------------------|
| Red Sindhi | RSi  | 39 | 0.68                    | 0.03      | 6.35 | 4.72                     |
| Tharparker | Thr  | 35 | 0.61                    | 0.03      | 5.20 | 3.95                     |
| Haryana    | Har  | 10 | 0.71                    | 0.03      | 5.15 | 5.01                     |
| Sahiwal    | Sah  | 10 | 0.60                    | 0.05      | 4.40 | 4.27                     |
| Desi       | Des  | 26 | 0.72                    | 0.02      | 6.55 | 5.11                     |
| Nellore    | Nel  | 27 | 0.65                    | 0.03      | 5.45 | 4.13                     |
| Ongole     | Ong  | 32 | 0.65                    | 0.03      | 5.05 | 4.12                     |

Exp Hz=Expected heterozygosity, MNA=Mean number of alleles per locus.

where  $P_E$  is the average frequency of the allele in the 11 European breeds and  $P_I$  is the average frequency in the seven Indian breeds assayed. The higher of these two values is denoted by the denominator. This SFD measure was combined with other criteria to define three categories of PAAs, each a subset of the previous. Initially, zebu alleles were defined as those that are at higher frequency in the Indian population than in the European population, and taurine alleles are the converse. Category I PAAs include both zebu and taurine alleles that have a frequency of greater than 5% in the respective parental population. Alleles with a frequency lower than this were removed from the analysis. Category II PAAs are the subset of category I, which are taurine alleles, that is, those found at higher frequency within *B. taurus* populations. Category III PAAs are a more stringent selection of taurine-associated alleles and are those found at a frequency of at least 20% in European populations and an SFD of greater than 40%. The average frequencies of categories II and III PAAs were calculated for each breed and these values were used as measures of *B. taurus* admixture.

#### Support for selection of PAAs (i) – The stepwise mutation model

The stepwise model predicts that ancestry informative alleles should cluster in size in an allele length frequency histogram as they are assumed to have evolved from an ancestral allele in one or other species. We therefore consider whether alleles designated as diagnostic of one or other ancestral group showed a tendency to cluster in size.

A total of 10 000 allele frequency distributions were simulated for each locus. Firstly, the number of zebu- and taurine-assigned alleles were taken as in the observed data using the category I criteria. Secondly, the order in which the alleles occurred on the length distribution was randomised. Every locus simulation was scored for allele intermingling as follows: An allele adjacency index (AAI) was calculated for each simulation by adding one when alleles designated to be from a single parental population were adjacent and subtracting one when alleles designated to be from different populations were adjacent. This process was repeated for each of the loci and the total score for every 20 loci simulation was calculated and is plotted as a histogram in Figure 3. The AAI for the observed data at the 20 loci was calculated and compared to the distribution of simulations. This analysis was repeated excluding any allele that did not have an SFD of at least 40% (data not shown).

#### Support for selection of PAAs (ii) – taurine alleles in the Near East

Allele frequencies for categories II and III taurine PAAs were calculated in seven Near Eastern cattle breeds using the MICROSATELLITE TOOLKIT (<http://oscar.gen.tcd.ie/~sdepark/ms-toolkit/index.html>) among which a cline of ancestry has been described previously (Loftus *et al*, 1999). The frequencies of both of these categories of alleles were averaged to give an estimate of the fraction of taurine ancestry in these breeds. A Spearman rank correlation was carried out where actual rank is the average frequency of taurine alleles and expected rank is based on latitude.

#### Genetic distance

Nei's standard genetic distance was calculated using MICROSAT (Takezaki and Nei, 1996; Minch *et al*, 1997). The distance between each European breed and each Near Eastern and Indian breed was calculated separately and subsequently averaged.

#### Structure analysis

The structure analysis program (Pritchard *et al*, 2000) uses allele frequency data to characterise parental populations, assign individuals to these populations, detect admixed individuals and estimate individual admixture. The program uses a model-based clustering method that employs a Markov Chain to estimate the posterior distribution ( $q$ ) of each individual's admixture co-efficient. The mean of this distribution has been denoted  $\hat{q}$  and represents an estimate of the amount of an individual's genome that is derived from one of the inferred parental populations. No prior information was used to characterise the parental populations or to assign individuals to those populations. We chose to infer two parental populations following the guidelines recommended by Pritchard *et al* (2000). A burn-in of 50 000 repetitions and a run length of 1 000 000 were used.

## Results

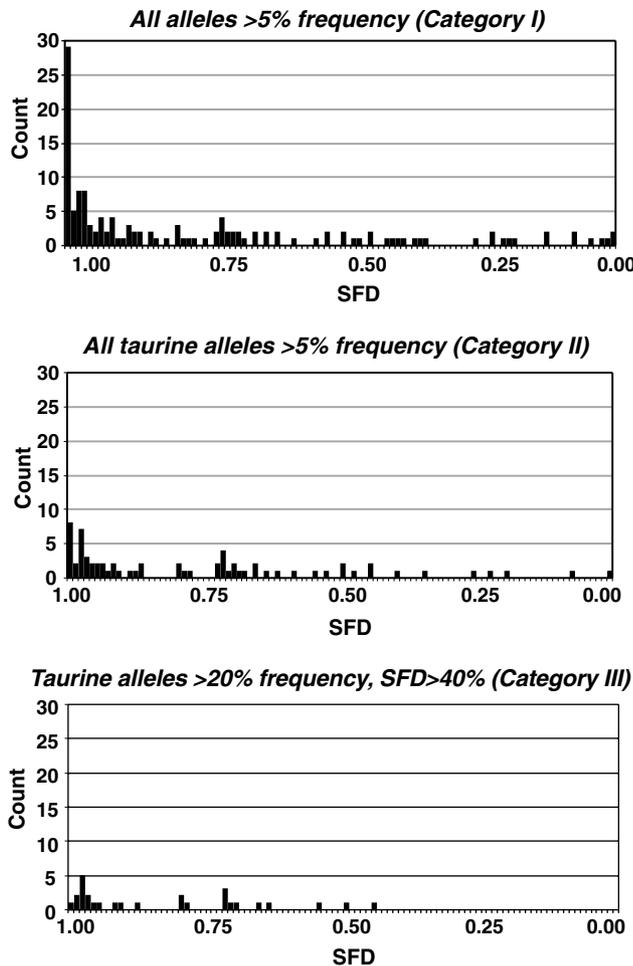
#### Diversity measures

Diversity indices for the samples examined in this study are given in Table 1. Within the South Asian breeds, neither gene diversity nor MNA showed a discernible geographical pattern. Gene diversity varied from 0.61 (Sahiwal) to 0.72 (Desi). In the five samples of comparable size, MNA values are lower in the two breeds of southern provenance, Ongole and Nellore than in Desi, Tharparkar, and Red Sindhi. These measures may reflect differences in historical population size but may also be a result of higher levels of taurine admixture in the more northern samples (see below).

#### Selection of PAAs

PAAs have been used in many studies to infer admixture proportion. It is a method that is especially useful when there are large allele frequency differences between the two parental populations. *B. indicus* and *B. taurus* cattle are sufficiently divergent such that there are many loci that show substantial allele frequency differences between the two species. These alleles have been used in previous admixture studies and produced results that have been supported by mitochondrial DNA and Y-chromosome studies. However, the accuracy of such analysis is dependent on correct assignment of the parental populations.

Here, we have selected alleles that are likely to be ancestry informative by picking alleles that show a high frequency differential between a pooled population of European taurine cattle ( $n=383$ ) and the pooled data from each Indian sample ( $n=178$ ). A pooled Indian sample was used to avoid designating one of the Indian breeds as a putative parental population. Figure 2 shows histograms of SFDs. These have been calculated for the



**Figure 2** Histogram of scaled frequency differential (SFD) values for three groups of alleles typed in the European and Indian populations at 20 microsatellite loci. The SFD is a function of the allele frequency differential between the European sample and the entire Indian sample (see Materials and methods section). The first histogram (I) shows all alleles that occur at a frequency of at least 5% in either population. Many alleles in this group have an SFD of 1 as they are unique to one population. Taurine alleles are shown in the second and third histograms. These occur at a higher frequency in Europe than India and can be used to detect taurine admixture into Indian zebu cattle breeds. Each category is a subset of the previous one.

following three allele categories:

- (I) All alleles that occur at a frequency of 5% or greater in either the Indian or European pool.
- (II) Taurine alleles that have a frequency of greater than 5%.
- (III) Taurine alleles that have a frequency of greater than 20% and an SFD of greater than 40%.

Taurine alleles are those that have a greater frequency in European breeds than Indian breeds. The category I histogram demonstrates a predominance of alleles that have a high frequency differential between the two species. The alleles in the category III histogram that occur at 20% or more in the European population and have an SFD value greater than 40% are likely to be ancestry informative and have been utilised in the following analyses (Figure 2). An SFD value of 40% excludes few taurine alleles and similar levels of

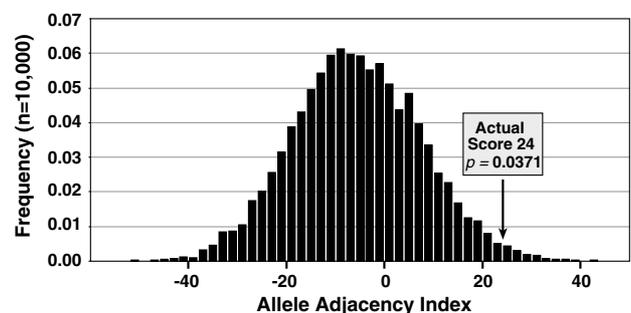
frequency differential have been found to be useful in other studies (McKeigue, 1998; Parra *et al*, 2001).

#### Support for PAA selection (i)

The stepwise model predicts that microsatellite alleles will mutate in discrete increments of repeat size (Shriver *et al*, 1993). If we have selected biologically meaningful PAAs, we would expect alleles that are diagnostic of one lineage to cluster in size. We wanted to ascertain whether this was the case for the loci in the present study as it acts as an independent control for our allele selection method. Therefore, 10 000 sets of 20 loci were simulated with allele order on the allele length axis randomised and given an adjacency score. This AAI is a measure of how often alleles from the same parental population were found to be adjacent. The AAI null distribution and the score for the actual data are shown in Figure 3. This score is higher than 95% of the simulated scores indicating that the alleles from one lineage are significantly less mixed than would be expected by chance ( $P = 0.0371$ ). A more stringent allele designation method where alleles are required to have an SFD of at least 40% produces a highly significant result ( $P = 0.0045$ ). This indicates that ancestry-informative alleles from one species have a significant nonrandom tendency towards similarity in size and thus supports common ancestral origins for selected taurine markers.

#### Confirmation of PAA selection (ii)

To further test our selection method, the average frequencies of categories II and III taurine PAAs in Near Eastern cattle populations were calculated. These populations were not used in the allele selection process and have been previously analysed by Loftus *et al* (1999). The average taurine allele frequency for each breed was compared to breed longitude using a Spearman rank correlation. The  $P$ -value for this test is less than 0.001, supporting a decline in taurine influence (or increase in zebu influence) from west to east through the Near East region, as would be predicted from geography and prior analysis (Loftus *et al*, 1999).



**Figure 3** Histogram of allele adjacency index (AAI) for 20 simulated microsatellite loci. A total of 10 000 sets of 20 loci were simulated in which the number of taurine and zebu alleles in each locus was held equal to the observed data but where the order in which they occur in the allele length distribution was randomised. The number of zebu and taurine alleles at the 20 simulated loci was determined by designating the alleles at the 20 study loci as either zebu or taurine using the category I allele criteria. An AAI was calculated for each simulation and these scores were plotted as a histogram which is unimodal. The score for the real data is shown by an arrow.

Admixture is known to truncate genetic distance. Hence, the average of Nei's standard genetic distance from the seven European breeds was taken for each Near Eastern population and plotted against PAAs frequency. The results for the two taurine PAAs categories are shown in Figure 4a. There is a negative correlation between the amount of taurine influence and the genetic distance from Europe ( $r^2(\text{adj}) = 0.69$ ,  $P < 0.01$  for category III PAAs).

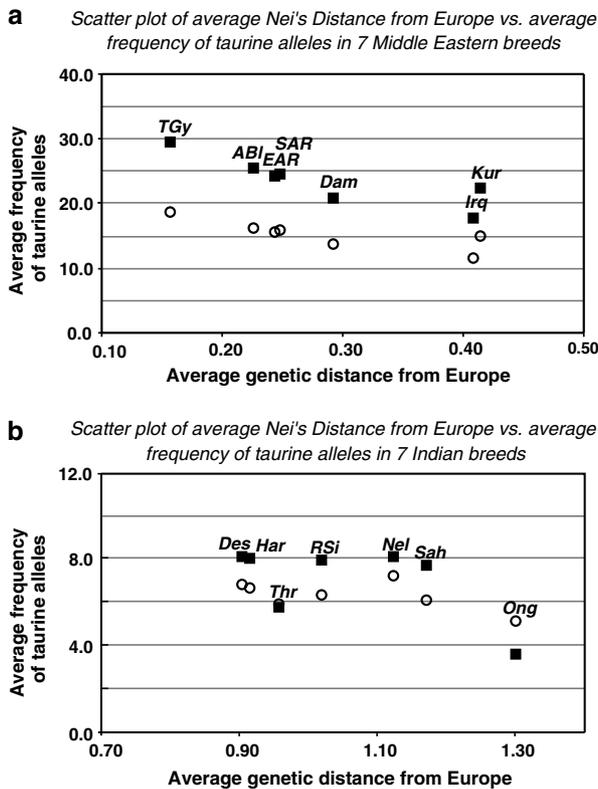
#### Application of PAAs to India

We calculated the average frequencies of our two categories of taurine PAAs in seven Indian cattle breeds. The southern-most Ongole breed has the lowest average frequency of both classes of taurine alleles. The average taurine allele frequency for each breed was used in a Spearman rank correlation using geographical latitude as the ranking parameter. This test does not support any simple cline throughout the data. The average of Nei's standard genetic distance from the seven European

breeds was plotted against the average taurine allele frequency of categories II and III alleles (Figure 4b). From both these data treatments, the Ongole breed appears to be the least introgressed of the Indian breeds.

#### STRUCTURE analysis

The STRUCTURE-integrated Bayesian approach both infers parental population allele frequencies and estimates the mean of the posterior distribution of each individual's admixture coefficient ( $\hat{q}$ ) (Pritchard *et al*, 2000). We chose to assume two parental populations for the admixture analysis. This supposition is supported by the inferred allele frequencies for the two parental populations, between which the average SFD is 0.7. The posterior distributions of admixture proportions for all individuals are not uniform and some have very narrow distributions that do not span either 0 or 1. Clearly, one of the two inferred populations is characterised by European *B. taurus* individuals and the other by Indian *B. indicus* cattle. These have  $\hat{q}$  values between 0.9 and 1.0 (taurine), and 0.0 and 0.1 (zebu). Beaumont *et al* (2001) used these values of  $\hat{q}$  to differentiate between parental ( $\hat{q} = 0-0.1$  and 0.9-1.0) and hybrid populations ( $\hat{q} = 0.1-0.9$ ). We can argue that this output reflects the ancestral divergence between *B. taurus* and *B. indicus*. The values of  $\hat{q}$  for each individual have been plotted according to breed in the box plot in Figure 5. The breeds with the highest and lowest median  $\hat{q}$  values are the Desi (0.022) and the Ongole (0.009) breeds, respectively. Values of  $\hat{q}$  range from only 0.006 to 0.03 in Ongole samples, while  $\hat{q}$  values from 0.006 to 0.552 are observed in the Red Sindhi breed of Northern India.



**Figure 4** Scatter plots of average of Nei's standard genetic distance from Europe vs the average frequency of categories II and III taurine alleles. Nei's genetic distance was calculated between each of the seven (a) Near Eastern and (b) Indian breeds and each European breed. The distance to European *Bos taurus* was averaged for each of the Near Eastern and Indian breeds. This value was plotted against the average frequency of categories II and III alleles. The circles represent the category II alleles and the squares the category III alleles. The category II allele criteria require alleles to occur at a minimum frequency of 5% in Europe, while the category III alleles require a European frequency of 20%. The first plot shows a cline of taurine influence from west to east in the Near Eastern cattle breeds. The regression for this plot is significant ( $P < 0.01$ ). The second plot shows the Indian breeds. The regression for this line is not significant but it is clear that Ongole has the lowest frequency of categories II and III taurine alleles.

#### Discussion

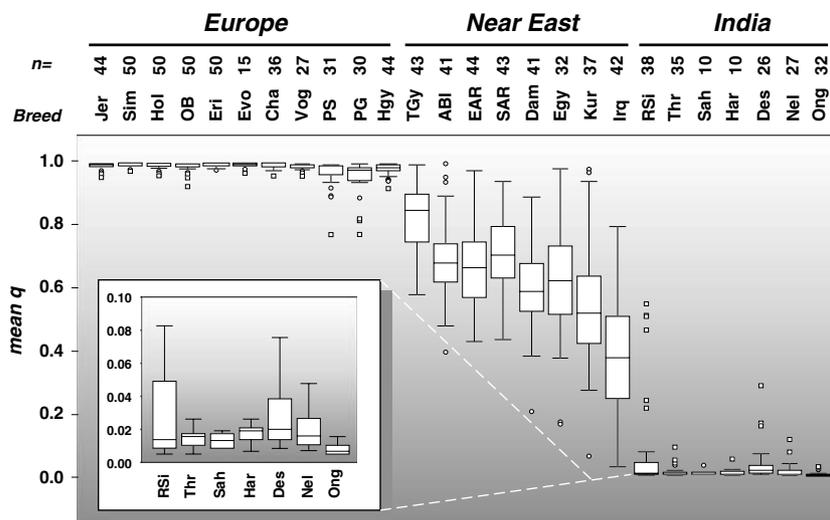
##### Patterns of genetic variability in the breeds surveyed

An older source population will typically display greater genetic diversity than a population derived more recently from it (Jorde *et al*, 1998). However, within the South Asian breeds, neither gene diversity (heterozygosity) nor MNA showed a discernible geographical pattern that could point to a likely region of origin (Table 1). In the five samples of comparable size, MNA values are lower in the two breeds of southern provenance, Ongole and Nellore than in Desi, Tharparkar, and Red Sindhi.

##### Introgression of *B. taurus*

Microsatellite allele size, as well as frequency, is a source of information about ancestral relationship. One feature of this property in cattle is that with a proportion of markers it seems that *B. indicus* and *B. taurus* populations display separate allele length distributions. Some of us have argued previously that through identification of allele classes, which were close to fixation in pure zebu, and absent or near-absent in breeds without their ancestral influence, the existence of geographical clines stretching from Europe to India can be shown (MacHugh *et al*, 1997; Loftus *et al*, 1999). Here, we take a symmetrical approach and estimate sets of alleles in loci that comprise the ancestral *B. taurus* distribution.

A potential obstacle to this approach is selection of the parental population. The selection of private alleles could



**Figure 5** Box plot of mean individual admixture coefficient ( $\hat{q}$ ) for each individual grouped by breed. Two parental populations were inferred by the STRUCTURE program without using prior sample information. One of these populations is characterised by European taurine and the mean of the posterior distribution of likely ancestry from this population was plotted for each individual. The hybrid status of the Near Eastern breeds is demonstrated by their position between Europe and India and the wide data range for each breed. The scale of the box plot makes it difficult to ascertain any difference between the seven Indian breeds, but if this section of the box plot is magnified (see the inset) it is clear that the Ongole samples have the lowest values of  $\hat{q}$ . The box represents the interquartile range that contains 50% of the values and line across the box indicates the median. The whiskers are the lines that extend from the box to the highest and lowest values, excluding outliers and extreme values. Outliers, which are represented by circles, are values that are 1.5–3 box lengths from the upper or lower edge of the box. Extreme values, which are greater than 3 box lengths, from the upper or lower edge are represented by squares.

be biased by the incorrect assignment of parental or hybrid status. The difficulties in estimating population and especially individual admixture when gene frequencies in the parental population before the admixture event are unknown have been discussed previously (Bertorelle and Excoffier, 1998; Estoup *et al*, 1999). Incorrect assignment of parental populations may lead to an overestimation of levels of admixture. In many cases, either temporal or spatial phenomena will mean that it is impossible to sample from the original parental populations that contributed to the admixture event. The issue of selecting ancestry-informative alleles is further confounded by the fact that some parental populations may have overlapping allele frequency distributions and there may be very limited numbers of private alleles that are found in only one or the other contributing population. In cases such as this, researchers are forced to select a reasonable frequency differential to characterise PAAs (Parra *et al*, 1998; Pfaff *et al*, 2001).

In the current study, we used three alternative methods to overcome these difficulties. An SFD between large pooled breed groups of alternately *B. taurus* and *B. indicus* cattle was used to assign alleles conservatively to parental populations. An examination of these illustrates relative influences of parental populations in hybrid populations. We also calculate the genetic distance of each of the seven Indian breeds from all European breeds using Nei's method. Finally, the STRUCTURE program, which does not require parental populations to be prespecified and which uses a Bayesian approach to define parental populations and to assess individuals, was employed.

The discriminating power of PAAs is implied when all alleles with a frequency greater than 5% are examined and show a high proportion of either taurine- or zebu-associated alleles (Figure 2, Category 1). Also, alleles at

any locus that are associated with one subspecies are more likely to be similar in repeat number; implying a genealogical relationship that supports their utility as indicators of ancestry (Figure 3).

The average frequencies of categories II and III alleles were calculated for seven Near Eastern breeds and plotted against Nei's standard genetic distance from Europe. There is a significant correlation between the two measures and a Spearman rank correlation using longitude as a ranking variable supports a gradation of average taurine allele frequency from west to east with a *P*-value of less than 0.001. This supports previous findings about the relationship between these breeds and is a reflection of the former herding of *B. indicus* alongside *B. taurus* in places such as Mesopotamia and Iran (Loftus *et al*, 1999). As these populations were not used in the initial allele designation, this pattern provides an independent test for whether the alleles are ancestry informative.

When we examined the frequency of these alleles in Indian breeds, *B. taurus* alleles were found in each population surveyed. While it is difficult to differentiate between the amounts of taurine influence in the Indian breeds, the PAA analysis suggests that the southern Ongole breed may have experienced the least taurine influence (Figure 4b).

The STRUCTURE program confirms the presence of *B. taurus* influence in the Indian subcontinent. The algorithm presumes that the original ancestral parental populations were in Hardy–Weinberg equilibrium, with the markers used in complete linkage equilibrium, and that the two parental populations contributed alleles to the hybrid population as independent draws. This model fairly accurately represents the original domesticated cattle populations of the India and the Near East, as it seems likely that they were separated for millennia prior

to any contact. STRUCTURE has the important advantage of modelling the ancestral parental population allele frequency distributions without *a priori* selection of these populations. It also allows admixed individuals to be included in the process of characterising these ancestral distributions. This approach also selects the Ongole as having the lowest likelihood of taurine influence (Figure 5). There is also a highly significant regression between the mean value of  $\hat{q}$  for Indian breeds and their average genetic distance from Europe ( $P < 0.0001$ ,  $r^2(\text{adj}) = 0.898$ ). Therefore, the data indicate some ancestral influence from taurine cattle, probably of Near Eastern origin. This is also suggested by the fact that highest values of individual admixture are in the Red Sindhi breed, which is geographically closest to the Near East.

The majority of Near Eastern individuals have posterior distributions with means between 0.1 and 0.9; a result of their documented hybrid status (Loftus *et al*, 1999). Both Beaumont *et al* (2001) and Hansen *et al* (2001) found that most hybrids with intermediate values of  $\hat{q}$  had 90% posterior intervals (PI) that ranged from 0.0 to 1, making it difficult to reliably identify admixed individuals. By contrast,  $\hat{q}$  here is between 0.1 and 0.9 for 33.6% of the sample in the present study and of these hybrid individuals 67% have 90% PI that are between 0.1 and 0.9. As the two previously mentioned studies used only nine microsatellite loci and we have included 20 in the current study, the suggestion that an increased number of markers will enhance resolution (Hansen *et al*, 2001) is supported by our results.

The extent of taurine influence in some of the Indian samples is somewhat surprising given the distinctly zebu morphology of all of the breeds surveyed. However, this is part of a broad geographic pattern of introgression with a cline in admixture stretching through the fringes of Europe, Anatolia, and the Middle East towards India. The presence of taurine alleles in India may be explained by either an ancient or more recent introduction of *B. taurus* into the Indian subcontinent. It is worth noting that there may have been attempts to improve some modern Indian cattle by very recent importation of *B. taurus* individuals (Feliu, 1995). However, in the current study, the unimproved, nondescript Desi cattle act as a control for this possible recent influence.

Locally domesticated zebu cattle were a prominent component of the Harappan civilisation of the Indus valley in the third millennium BC, as indicated by faunal and artistic evidence (Zeuner, 1963; Allchin, 1969; Meadow, 1996). By this period, however, some humpless taurine cattle were also present, as they are depicted on some Harappan stamp seals. In addition, evidence, both from ancient art as well as some bone assemblages, indicates that zebu were present in Sistan, Iran, and further west in Mesopotamia and even Kamid el-Loz in Lebanon during the Bronze Age, that is, 3000–1000 BC (Clason, 1978; Meadow, 1987; Bököyni, 1990, 1997). This evidence for the spread of zebu westwards to the Near East and the spread of taurine cattle eastwards to the Indus valley supports that the introgression between taurine and zebu cattle indicated by the genetic data could have begun by the third millennium BC and has continued in varying degrees ever since. Archaeological evidence for long-distance trade indicates that there has been continuous contact between the Indus region and

Mesopotamia since the fourth millennium BC (Allchin and Allchin, 1982; Lamberg-Karlovsky, 1989; Chakrabarti, 1990).

It was also through this same region (Baluchistan) that a series of folk movements of peoples from the north-western periphery of South Asia are postulated to have occurred during the course of and after the Harappan civilisation (Allchin and Allchin, 1982). The scale of these migrations, and whether or not they are to be connected with the diffusion of Indo-European languages as traditionally assumed (Parpola, 1994), remains highly controversial (Shaffer, 1984; Renfrew, 1987; Chakrabarti, 1997). Nevertheless, such human movements, presumably of agropastoral groups, would have contributed to the further introgression of taurine genes into zebu populations.

At this period, or perhaps in preceding millennia, flax and a number of pulses, including peas, lentils, grasspea, and chickpea, spread from west to east. During the third millennium some subsistence species spread westwards, including sesame (Charles, 1993), the chicken (Bököyni, 1990), and water buffaloes (Zeuner, 1963). Therefore, the dispersal of *B. taurus* cattle into South Asia, and the counterflow of some *B. indicus* into Mesopotamia and the Levant can be seen as part of a larger pattern of the prehistoric transfer of subsistence species between these two primary centres of agricultural innovation (Fuller, 2001).

## Acknowledgements

We acknowledge the assistance of all those who participated in the collection of samples. Thanks also to SDE Park for his help with data analysis. P Kumar was funded by FAO under an Indian Fellowship UTF/IND/162. AR Freeman is supported by a Wellcome Studentship in Biodiversity number 054275/Z/98/Z. RT Loftus was supported by the European Commission DG XII Biotechnology Programme, Contract No. BIO4 CT961189. We gratefully acknowledge two anonymous reviewers for their helpful comments.

## References

- Allchin FR (1969). Early domestic animals in India and Pakistan. In: Ucko PJ, Dimbleby GW (eds) *The Domestication and Exploitation of Plants and Animals*, Duckworth: London. pp 317–322.
- Allchin B, Allchin FR (1982). *The Rise of Civilization in India and Pakistan*. Cambridge University Press: Cambridge.
- Beaumont M, Barrett EM, Gotelli D, Kitchener AC, Daniels MJ, Pritchard JK *et al* (2001). Genetic diversity and introgression in the Scottish wildcat. *Mol Ecol* 10: 319–336.
- Bertorelle G, Excoffier L (1998). Inferring admixture proportions from molecular data. *Mol Biol Evol* 15: 1298–1311
- Bököyni S (1990). *Kamid el-Loz. 12. Tierhaltung und Jagd*, Saarbrücker Beiträge zur Altertumskunde band 42. Dr. Rudolf Habelt GmbH: Bonn.
- Bököyni S. (1997). Zebu and Indian wild cattle. *Anthropozoologica* 2526: 647–654.
- Bradley DG, MacHugh DE, Loftus RT, Sow RS, Hoste CH, Cunningham EP (1994). Zebu-taurine variation in Y-chromosomal DNA – a sensitive assay for genetic introgression in West-African trypanotolerant cattle populations. *Anim Genet* 25: 7–12.
- Chakrabarti DK (1990). *The External Trade of the Indus Civilization*. Munshiram Manoharlal: New Delhi.

- Chakrabarti DK (1997). *Colonial Indology. Sociopolitics of the Ancient Indian Past*. Munshiram Manoharlal: New Delhi.
- Charles MP (1993). Botanical remains. In: Green A (ed) *Abu Salabikh Excavations. The 6G Ash-tip and its Contents: Cultic and Administrative Discard from the Temple*, British School of Archaeology in Iraq: London. Vol 4, pp 203–207.
- Clason AT (1978). Late Bronze Age-Iron Age Zebu in Jordan? *J Archaeol Sci* 5: 91–93.
- Collins-Schramm HE, Phillips CM, Operario DJ, Lee JS, Weber JL, Hanson RL et al (2002). Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. *Am J Hum Genet* 70: 737–750.
- Edwards CJ, Dolf G, Looft C, Loftus RT, Bradley DG (2000). Relationships between the endangered Pustertaler-Sprinzen and three related European cattle breeds as analysed with 20 microsatellite loci. *Anim Genet* 31: 329–332.
- Estoup A, Cornuet JM, Rousset F, Guyomard R (1999). Juxtaposed microsatellite systems as diagnostic markers for admixture: theoretical aspects. *Mol Biol Evol* 16: 898–908.
- Felius M. (1995). *Cattle Breeds: an Encyclopaedia*. Misset: Doetinchem, Netherlands.
- Fuller DQ (2001). Fifty years of archaeobotanical studies in India: laying a solid foundation. In: Settar S, Korisettar R (eds) *Indian Archaeology in Retrospect, Archaeology and Interactive Disciplines*, Publications of the Indian Council for Historical Research, Manohar: New Delhi. Vol 3, pp 247–363.
- Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JE (2002). African pastoralism: genetic imprints of origins and migrations. *Science* 296: 336–339.
- Hanotte O, Tawah CL, Bradley DG, Okomo M, Verjee Y, Ochieng J et al (2000). Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds. *Mol Ecol* 9: 387–396.
- Hansen MM, Neilsen EE, Bekkevold D, Mensberg KL (2001). Admixture analysis and stocking impact assessment in brown trout (*Salmo trutta*), estimated with incomplete baseline data. *Can J Fish Aquat Sci* 58: 1853–1860.
- Jorde LB, Bamshad M, Rogers AR (1998). Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* 20: 126–136.
- Lamberg-Karlovsky CC (1989). Mesopotamia, Central Asia and the Indus Valley: so the kings were killed. In: Lamberg-Karlovsky CC (ed) *Archaeological Thought in America*, Cambridge University Press: Cambridge. pp 241–267.
- Loftus RT, Ertugrul O, Harba AH, El-Barody MAA, MacHugh DE, Bradley DG (1999). A microsatellite survey of cattle from a centre of origin: the Near East. *Mol Ecol* 8: 2015–2022.
- Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham EP (1994). Evidence for two independent domestications of cattle. *Proc Natl Acad Sci USA* 91: 2757–2761.
- MacHugh DE, Shriver MD, Loftus RT, Cunningham EP, Bradley DG (1997). Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* 146: 1071–1086.
- McKeigue PM (1997). Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. *Am J Hum Genet* 60: 188–196.
- McKeigue PM (1998). Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am J Hum Genet* 63: 241–251.
- Meadow RH (1987). Faunal exploitation patterns in Eastern Iran and Baluchistan: a review of recent investigations. In: Gnoli G, Lanciotti L (eds) *Orientalia Iosephi Tucci Memoriae Dicata. Serie Orientale Roma LVI, 2*, Istituto Italiano per il Medio ed Esterno Oriente: Rome. pp 881–916.
- Meadow RH (1996). The origins and spread of agriculture and pastoralism in northwestern South Asia. In: Harris, OR (ed) *The Origins and Spread of Agriculture and Pastoralism in Eurasia*, UCL Press Ltd: London. pp 390–412.
- Minch E, Ruiz-Linares A, Goldstein DB, Feldman MW, Cavalli-Sforza LL (1997). *MICROSAT: a computer program for calculating various statistics on Microsatellite Allele Data*. Version 1.5d. Department of Genetics, Stanford University, Stanford.
- Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.
- Parpola A (1994). *Deciphering the Indus Script*. Cambridge University Press: Cambridge.
- Parra EJ, Kittles RA, Argyropoulos G, Pfaff CL, Hiester K, Bonilla C et al (2001). Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am J Phys Anthropol* 114: 18–29.
- Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R et al (1998). Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63: 1839–1851.
- Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI et al (2001). Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am J Hum Genet* 68: 198–207.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Renfrew C (1987). *Archaeology and Language – The Puzzle of Indo-European Origins*. Cambridge University Press: Cambridge.
- Ritz LR, Glowatzki-Mullis M-L, MacHugh DE, Gaillard C (2000). Genetic diversity in the tribe Bovidi revealed with microsatellite polymorphisms. *Anim Genet* 31: 178–185.
- Shaffer JG (1984). The Indo-Aryan invasions: cultural myth and archaeological reality. In: Lukacs JR (ed) *The People of South Asia: The Biological Anthropology of India, Pakistan and Nepal*, Plenum: London. pp 77–90.
- Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993). VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* 134: 983–993.
- Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R et al (1997). Ethnic-affiliation estimation by use of population-specific DNA markers. *Am J Hum Genet* 60: 957–964.
- Takezaki N, Nei M (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144: 389–399.
- Wilson JF, Goldstein DB (2000). Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. *Am J Hum Genet* 67: 926–935.
- Zeuner FE (1963). The history of the domestication of cattle. In: Mourant AE, Zeuner FE (eds) *Man and Cattle*, Royal Anthropological Institute of Great Britain and Ireland, London. Vol 18, pp 9–20.