Microsatellite DNA Variation and the Evolution, Domestication and Phylogeography of Taurine and Zebu Cattle (*Bos taurus* and *Bos indicus*)

David E. MacHugh,* Mark D. Shriver,† Ronan T. Loftus,* Patrick Cunningham* and Daniel G. Bradley*

**Department of Genetics, Trinity College, Dublin 2, Ireland and †Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*

Manuscript received December 9, 1996
Accepted for publication March 26, 1997

ABSTRACT

Genetic variation at 20 microsatellite loci was surveyed to determine the evolutionary relationships and molecular biogeography of 20 different cattle populations from Africa, Europe and Asia. Phylogenetic reconstruction and multivariate analysis highlighted a marked distinction between humpless (taurine) and humped (zebu) cattle, providing strong support for a separate origin for domesticated zebu cattle. A molecular clock calculation using bison (*Bison* sp.) as an outgroup gave an estimated divergence time between the two subspecies of 610,000–850,000 years. Substantial differences in the distribution of alleles at 10 of these loci were observed between zebu and taurine cattle. These markers subsequently proved very useful for investigations of gene flow and admixture in African populations. When these data were considered in conjunction with previous mitochondrial and *Y* chromosomal studies, a distinctive male-mediated pattern of zebu genetic introgression was revealed. The introgression of zebu-specific alleles in African cattle afforded a high resolution perspective on the hybrid nature of African cattle populations and also suggested that certain West African populations of valuable disease-tolerant taurine cattle are under threat of genetic absorption by migrating zebu herds.

MICROSATELLITES or simple tandem repeat (STR) genetic markers have become the mainstay of genetic linkage mapping efforts. Recently, they have also been used to address questions concerning the genetic diversity and evolutionary history of various organisms. In particular, the genetic origins of human populations have been usefully explored using microsatellite polymorphisms (BOWCOCK *et al.* 1994; DEKA *et al.* 1995; GOLDSTEIN *et al.* 1995a). We have also previously used microsatellite polymorphisms to investigate genetic variation among European breeds of cattle (MACHUGH *et al.* 1994).

Microsatellite analysis of genetic diversity provides two distinct levels of information. In addition to allele frequency differences among populations, it also provides information about the cladistic relationships between alleles and groups of alleles on the basis of differences in allelic repeat length. Novel methods have therefore been developed to investigate the mutational dynamics of microsatellites and their application to population genetic problems. The stepwise mutation model (SMM), in which microsatellite allele length is treated as an incremental quantitative character, has been an important component of these new analytical methods (SHRIVER *et al.* 1993, 1995; GOLDSTEIN *et al.* 1995b).

Domesticated cattle are the major component of pastoral economies throughout the world. Through milk, they provide the bulk of the animal protein consumed by many human societies, and contribute other important commodities including meat, hides, traction and dung. The biological systematics and evolutionary relationships of cattle have always been highly contentious. In particular, the relationship between the two main types of cattle, humped zebu (*Bos indicus*) and humpless taurine (*Bos taurus*), has been an active area of research with various hypotheses having been proposed to account for the morphological and genetic differences observed between the two subspecies.

One school of thought asserts that domesticated cattle were first developed from a single wild ancestor, the aurochs (*Bos primigenius primigenius*) during the early Neolithic phase of the agricultural communities in the Near East (circa 8000–9000 BP). *Bos indicus* populations are then thought to have been produced at a later date through breeding and selection from *Bos taurus* cattle (EPSTEIN 1971; EPSTEIN and MASON 1984; PAYNE 1991). The alternative viewpoint contends that domesticated zebu populations were developed independently by a separate group of early pastoralists and that the candidate domestication centers were the Neolithic societies of Baluchistan in present-day Pakistan. The southern Asian subspecies of aurochs (*Bos primigenius namadicus*) would then be the most likely progenitor of domesticated zebu cattle. This interpretation is supported by recent archaeological and genetic evidence (MEADOW 1993; LOFTUS *et al.* 1994a; BRADLEY *et al.*

Corresponding author: David MacHugh, Genetics Department, Trinity College, Dublin 2, Ireland. E-mail: dmachugh@mail.tcd.ie

1996). In particular, through analysis of mitochondrial DNA (mtDNA) sequence variation in extant cattle populations, Loftus *et al.* (1994a) were able to demonstrate that taurine and zebu cattle probably diverged at least 210,000 years ago, well outside the range required for human-mediated development of zebu cattle from taurine progenitors.

The history and biogeography of cattle populations in Africa represent a complex interaction of ecological, genetic and anthropological factors. The original indigenous cattle of Africa are universally considered to have been exclusively taurine. These populations are thought to have arisen from migrations of early pastoralists from the Near East (EPSTEIN 1971; PAYNE 1991). Recent archaeological evidence has, however, questioned this viewpoint, suggesting that the African aurochs (*Bos primigenius opisthonomus*) may have been domesticated independently somewhere on the African continent (WENDORF and SCHILD 1994).

Although, zebu cattle are thought to have been first introduced into Africa about 4000 years ago, they only started to become widespread about 700 AD with the Arabic migrations into North and East Africa. At present, Africa represents a mosaic of cattle morphologies with zebu cattle and intermediate forms, often referred to as "sanga," predominating over most of the continent (PAYNE 1970; EPSTEIN 1971). It has previously been suggested that the genetic signature of zebu gene flow in Africa may represent an unusual pattern of male-mediated introgression (BRADLEY *et al.* 1994; LOFTUS *et al.* 1994a; BRADLEY *et al.* 1996). Taurine populations predominate only in subtropical regions of West Africa where an inherited resistance to trypanosomiasis (trypanotolerance) allows them to inhabit areas infested with tsetse flies (*Glossina* sp.). Zebu cattle possess no innate resistance to trypanosomiasis and have only started to penetrate these regions with the assistance of veterinary prophylaxis and the destruction of the tsetse habitat through widespread deforestation (LHOSTE 1991). These recent migrations of zebu cattle pose a serious threat to the genetic integrity of valuable trypanotolerant populations of taurine cattle in the southern areas of West and Central Africa.

To clarify the genetic relationships among the major groups of cattle and to characterize the extent and pattern of zebu genetic introgression in African populations, we have screened 20 different microsatellite loci in a large number of cattle representing 20 distinct populations from Africa, Europe and Asia. The patterns of genetic variation observed within and among populations should provide a large body of data to test hypotheses concerning the genetic affinities and origins of domesticated cattle, the dynamics of zebu genetic introgression in Africa and should also provide a method to assess the genetic integrity of threatened taurine populations in West Africa.

MATERIALS AND METHODS

Sample collection and DNA extraction: DNA samples were collected from 728 individual cattle representing 20 different populations originating from Africa, Asia and Europe. Samples were collected during a series of field sampling missions from artificial insemination (AI) stations, agricultural research institutions and village herds. Every attempt was made to ensure each population sample was representative. A strict, distributed sampling strategy was employed and herd book and breeding records were consulted while farmers and breeders were questioned about the origins and familial relationships of individual animals. The populations, their sample sizes and other relevant information are detailed in Table 1. In addition, a small number of samples were taken from three related species. These were American bison (*Bison bison*; $n = 2$), European bison (*Bison bonasus*; $n = 2$) and Banteng (*Bibos banteng*; $n = 2$). DNA was extracted from blood samples using standard procedures (SAMBROOK *et al.* 1989). Semen straws were used as a DNA source for some of the European breeds and these were processed according to a protocol described by ANDERSSON *et al.* (1986).

Microsatellite genotype analysis: A battery of 36 randomly chosen microsatellite loci were used from the array of public-domain markers available in the bovine genome mapping literature. These microsatellite markers were variants of the $(CA)_n$ class and were characterized and developed using standard methods. Eventually, 20 of the 36 markers that amplified robustly and repeatedly were used for full screens of the total sample panel. These 20 microsatellites are detailed in Table 2 with associated reaction conditions. PCR amplifications of individual loci were performed as follows. Reactions were performed using 96-well microtitre plates with 5–10 ng template DNA in 11 μl reaction volumes using 0.5 unit of *Taq* polymerase with reaction buffer comprising 50 mM KCl; 10 mM Tris-HCl pH 9.0; 1.0–2.5 mM $MgCl_2$ (see Table 2); 1% Triton X-100; 200 μM dATP, dGTP, dTTP; 10 μM dCTP; 0.3 μM of each primer was added, as was 0.5 μCi [α-^{32}P] dCTP. A 10 μl oil overlay was then added and amplifications were performed in a Hybaid OmniGene thermal cycler using a 4 min denaturation step at 94°, followed by 35 cycles of 45 sec at 93°, 45 sec at 54–63° (see Table 2), 45 sec at 72° and a final extension step at 72° for 4 min. Samples were then mixed with 10 μl formamide solution. After heat treatment at 93°, 3-μl aliquots of these mixtures were then loaded on 6% denaturing polyacrylamide sequencing gels. M13mp18 plasmid DNA was used to provide sequence ladders for initial allele size calibration. Samples with appropriate genotypes were then selected to provide unambiguous length standards for each microsatellite. The allelic data from each autoradiogram were entered manually into a computer database. Data for each gel was then printed out and double-checked against the original autoradiogram.

Basic data analysis: Allele frequencies were determined by direct counting. Allele frequency distributions for the 20 microsatellite loci are available via anonymous file-transfer at the following ftp address: acer.gen.tcd.ie/pub/cow_microsat/.

As discussed below, 10 out of the 20 microsatellites displayed alleles that were present at high frequency in Asian zebu breeds, intermediate frequencies in African zebu populations, low frequencies in African taurine populations and that were either absent or at low frequencies in European breeds. As an illustration of this phenomenon, Figure 1 shows the allelic distributions for two of these loci in pooled populations samples. These private (NEEL 1973) or zebu-diagnostic alleles were used to evaluate gene flow and genetic admixture in African individuals and populations (see below). Observed heterozygosity and unbiased estimates of gene diversity (expected heterozygosity) with associated standard errors were

TABLE 1

Characteristics of the 20 cattle populations surveyed

Classification	Breed	Geographical origin	Sample size	Comments
European taurine	Aberdeen Angus	Scotland	33	—
European taurine	Hereford	England	34	—
European taurine	Jersey	Channel Islands	34	—
European taurine	Kerry	Ireland	40	Endangered[a]
European taurine	Charolais	France	36	—
European taurine	Friesian	Netherlands	40	—
European taurine	Simmental	Switzerland	36	—
African taurine	N'Dama	Gambia	58	Trypanotolerant[b]
African taurine	N'Dama	Guinea	63	Trypanotolerant[b]
African taurine	N'Dama	Guinea Bissau	54	Trypanotolerant[b]
African taurine	N'Dama	Nigeria	19	Trypanotolerant[b]
African taurine	N'Dama	Senegal	48	Trypanotolerant[b]
African zebu	Butana	North Sudan	24	—
African zebu	Kenana	South Sudan	38	—
African zebu	Gobra	Senegal	59	—
African zebu	Maure	Mauritania	55	Sampled in Senegal[c]
African zebu	White Fulani	Nigeria	24	—
Asian zebu	Hariana	North India	10	—
Asian zebu	Sahiwal	Pakistan	13	—
Asian zebu	Tharparker	North India	10	—

[a] The Kerry cattle breed has been classified as endangered by the Food and Agriculture Organization of the United Nations (SCHERF 1995).

[b] N'Dama cattle, in common with many West African taurine populations, display an inherited tolerance or resistance to the effects of tsetse-borne trypanosomiasis.

[c] Because of a border dispute between Mauritania and Senegal, the Maure animals were sampled from a dwindling remnant population in Senegal.

computed for all locus/population combinations according to NEI (1987).

Genetic stratification: Three tests for deviations from Hardy-Weinberg Equilibrium (HWE) were employed. The first two methods were performed as described by HAMMOND *et al.* (1994) and DEKA *et al.* (1995). In the first method a chi-square test was used to evaluate the overall discordance of genotype frequencies at each locus/population combination. The second method used was based on a likelihood ratio (L) test criterion (G-statistic) that was used to contrast observed and expected genotype frequencies. The levels of significance for these two tests were determined empirically by shuffling (permutation) of alleles across individuals with 1000 replications (CHAKRABORTY *et al.* 1991; EDWARDS *et al.* 1992). The third test for deviations from HWE was performed using the GENEPOP package version 1.2 (RAYMOND and ROUSSET 1995a). This method uses an exact test of HWE with the addition that, for loci/population combinations with five or more alleles, a Markov chain algorithm is used to obtain an unbiased estimate of the exact probability of being wrong in rejecting HWE. In all cases, the Markov chain was set to 50,000 steps with 1000 steps of dememorization. The significance of the HWE probabilities for the three tests across loci and populations was determined using Fisher's method for combining probabilities with the following formula as discussed by RAYMOND and ROUSSET (1995b).

$$\chi^2_{TOT} = -2\left(\sum_{j=1}^{r} \ln P_j\right),$$

where r is the number of loci when pooling across breeds or the number of breeds when pooling across loci (in either case $r = 20$), P_j is the P value for the jth locus or breed. This statistic follows a chi-square distribution with $2 \times r$ d.f.

Genetic differentiation: RAYMOND and ROUSSET (1995b) have proposed a powerful method to assess population differentiation based on whether allelic composition is independent of population assignment. This statistical test is based again on analysis of contingency tables using a Markov chain procedure to derive an unbiased estimate of the exact probability in being wrong in rejecting the null hypothesis (H_0); *i.e.*, allelic composition is independent of population assignment (no differentiation). This test was performed for all 190 pairwise interpopulation comparisons on contingency tables containing data from each of the 20 microsatellite loci. The GENEPOP package was employed with the same Markov chain parameters detailed above.

Genetic distance computations and phylogenetic reconstruction: The D_{SW} "stepwise-weighted genetic distance" was computed according to SHRIVER *et al.* (1995), using a modified version of the DISPAN computer package (DSW, T. OTA, Center for Human Genetics, Boston University). SHRIVER *et al.* (1995) have shown that the D_{SW} measure is an appropriate measure for microsatellite loci that are presumed to evolve in a stepwise or near-stepwise fashion (SHRIVER *et al.* 1993; VALDES *et al.* 1993). Neighbor-joining (N-J) trees (SAITOU and NEI 1987) were constructed from D_{SW} distances and the reliability of the trees obtained were examined by a bootstrap test with 1000 replicate resamplings of loci with replacement (FELSENSTEIN 1985).

Multivariate statistical analysis: To condense the genetic variation revealed with the panel of 20 microsatellites, principal components analysis (PCA) was performed according to the procedure described by CAVALLI-SFORZA *et al.* (1994). PCA involves a linear transformation of the observed allele frequencies (in geometrical terms a rotation of the coordinate axes) where the coefficients are chosen so as to maximize the variation of the transformed data measured along each new

TABLE 2

Chromosomal locations primer sequences and experimental parameters for 20 microsatellite markers

Locus[a]	Chromosome	Primer sequences (5′-3′)	Tm (deg)	$MgCl_2$ (mM)	Size range (bp)
ETH152 (D5S1)	*5*	TACTCGTAGGGCAGGCTGCCTG GAGACCTCAGGGTTGGTGATCAG	59	1.0	191–211
ETH225 (D9S1)	*9*	GATCACCTTGCCACTATTTCCT ACATGACAGCCAGCTGCTACT	57	1.5	140–160
HEL1 (D15S10)	*15*	AGTCCATGGGATTGAAAGAGTTGG CTTTTATTCAACAGCTATTTAACAAGG	55	1.0	101–117
HRH1	*22*	GGCTTCAACTCACTGTAACACATT TTCTTCAAGTATCACCTCTGTGGCC	55	2.5	180–190
ILSTS001 (D7S13)	*7*	GGTGCTGTTATCTAGAATTTGG GGAGTCATACACAACTGAGC	58	1.0	77–97
ILSTS005 (D10S25)	*10*	GGAAGCAATGAAATCTATAGCC TGTTCTGTGAGTTTGTAAGC	55	2.0	181–193
ILSTS014 (D19S10)	*19*	CTGACTATGGTGATAATCCC TCTTTTCCCTTTCCTTCCCC	58	2.0	128–134
OCAM	*25*	CCTGACTATAATGTACAGATCCCTC GCAGAATGACTAGGAAGGATGGCA	57	1.5	178–190
RASA	*7q24-qter*	CCCTTCCGCTTTAGTGCAGCCAG GGGCCACAGCCCAGGATCGGGAGC	63	1.5	182–196
TGLA48 (D7S26)	*7*	AAATGTTTTATCTTGACTACTAAGC ACATGACTCTGCCATAGAGCAT	57	1.5	73–79
BM2113 (D2S26)	*2*	GCTGCCTTCTACCAAATACCC CTTCCTGAGAGAAGCAACACC	55	1.0	123–147
BoLA DR2B	*23*	AGGCAGCGCCGAGGTGAGCGA TCCAACACTCACCTGGACGTAGC	60	1.5	144–152
BoLA DRP1	*23*	ATGGTGCAGCAGCAAGGTGAGCA GGGACTCAGTCTCTCTATCTCTTTG	55	1.0	118–142
BTMICROS	Unassigned	CTAGAAGATTTAGAAAACTTGCC ATAGCAAGACATATCTCCATTCC	54	1.5	141–187
ETH131 (D21S4)	*21*	GTGGACTATAGACCATAAGGTC GCTGTGATGGTCTACGAATGA	55	1.5	138–168
HBB	*15q22-q27*	GATATAAAAAAGAAGACCCAGTAG TACCTGAGTCATATGTAATATTCC	54	2.0	98–120
HEL5 (D21S15)	*21*	GGTAATGGTTTTCAGACGTTAGTG GTAGCAGGATCACTTGTTAGGG	54	1.5	161–181
PRL	*23*	GGAAAGTGAACATGACTGTCTAG GCCCTCTCTTCTACAATGAACAC	60	1.5	158–164
RBP3	*28*	TGTATGATCACCTTCTATGCTTC GCTTTAGGTAATCATCAGATAGC	55	1.5	141–153
TGLA116 (D4S4)	*4*	GCACAGTAAGAGTGATGGCAGA TGGAGAAGATTTGGCTGTGTACCCA	57	1.5	79–85

[a] Bovine genome map codes are shown in parentheses (EGGEN and FRIES 1995).

coordinate axis (PC). The first three principal components (PCs), by definition, are the most informative and these were plotted on a three-dimensional scatter diagram for all 20 populations.

Analysis of African gene flow and genetic admixture: Two methods were used to estimate Asian zebu introgression in African cattle populations. As mentioned above, 10 of the 20 microsatellites displayed zebu-diagnostic alleles (see Figure 1). The criteria used to determine whether an allele was zebu-diagnostic was presence at high frequency in the Asian zebu populations and absence in the N'Dama population sampled in Guinea. Guinean N'Dama cattle are considered to be a completely pure African taurine gene pool and previous analysis of zebu *Y*-chromosome polymorphisms has indicated that they have not been subject to zebu genetic introgression (BRADLEY *et al.* 1994). In total, 18 alleles from the 10 diagnostic loci were classified as zebu-diagnostic. The frequencies of these alleles or groups of alleles at a particular locus were averaged to give an estimate of the frequency of zebu-specific alleles in each of the African populations. The zebu-specific allele for one of these loci (a 182-bp OCAM allele) was present in the Guinean N'Dama at a very low frequency (1.6%). However, its distribution in India and across the rest of Africa would suggest that it is a genuine zebu-specific allele and it was considered as such for the purposes of this survey. The presence of this allele in Guinean N'Dama may therefore represent residual low-level zebu introgression or perhaps allelic homoplasy.

Estimation of genetic admixture proportions from all allele frequencies: A second, more sophisticated approach was used to estimate direct zebu admixture proportions from the total data set. This method has been developed by CHAKRABORTY (1975, 1985, 1986) and uses the concept of the gene identity coefficient, the probability that two genes chosen at random (from one or more populations) are identical in state (NEI 1972). The underlying rationale to this method is that genetic simi-

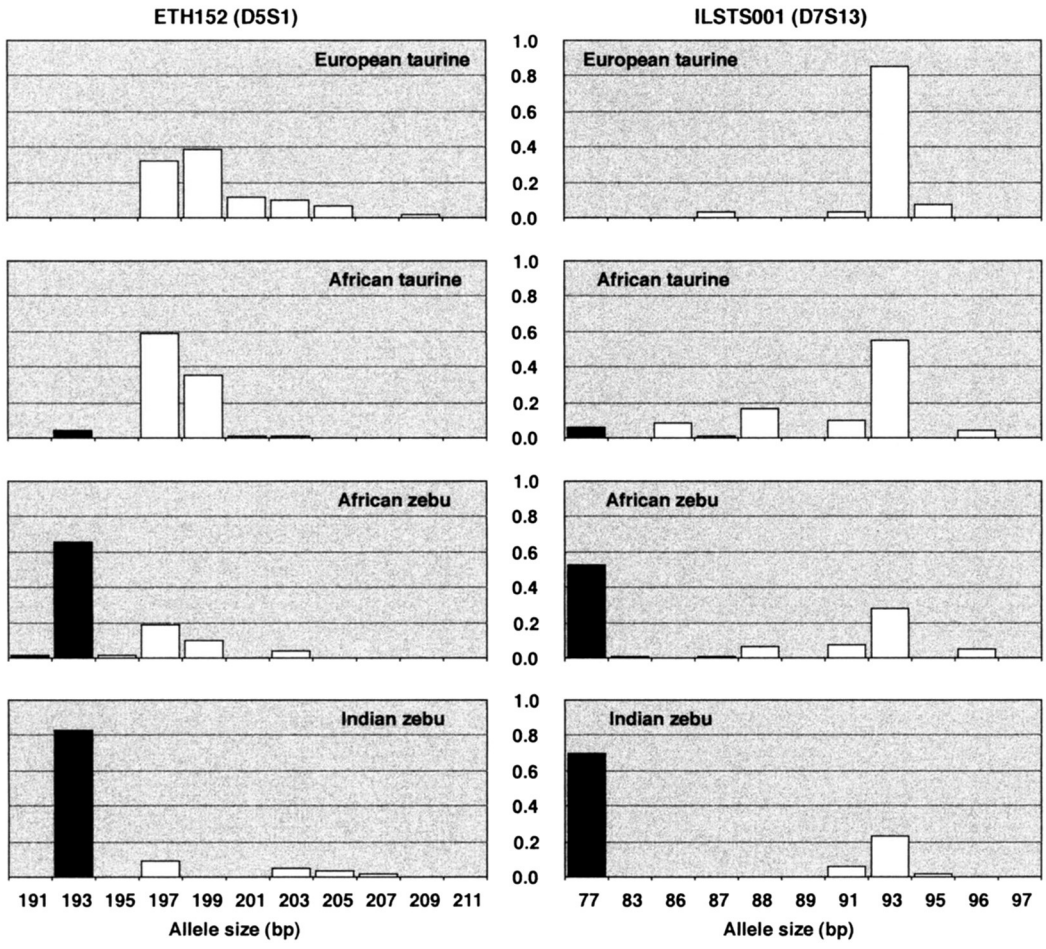

FIGURE 1.—Allele frequency distributions for two zebu-diagnostic loci. The pooled samples are as follows: (1) seven European taurine breeds, (2) five African N'Dama taurine populations, (3) five African zebu breeds and (4) three Indian zebu breeds. The diagnostic alleles for each locus are indicated by a solid black color. The diagnostic alleles are 191/193 bp for the ETH152 microsatellite and 77 bp for the ILSTS001 locus. The eight other zebu-diagnostic loci are as follows: ETH225, HEL1, HRH1, ILSTS005, ILSTS014, OCAM, RASA and TGLA48.

larity between populations can be expressed as a simple linear function of admixture proportions. This method requires that data be available from two putative source or that parental populations represent the original populations that produced the dihybrid populations of interest. An example of this is in studies of human admixture where present-day West African and Caucasian populations are used to represent the original source populations for investigations of genetic admixture in African-American populations (WORKMAN *et al.* 1963; CHAKRABORTY *et al.* 1992). In a similar fashion, the Indian zebu cattle breeds were pooled and used to represent a putative Asian zebu parental population and the Guinean N'Dama were used to represent a putative African taurine source population. The computations were performed using a computer program ADMIX (R. CHAKRABORTY, Genetics Center, University of Texas Health Science Center at Houston). This program uses a vector-matrix approach to produce a weighted least-square solution for each individual admixture proportion with an associated standard error (CHAKRABORTY 1985, 1986). It also produces correlation coefficients for the weighted least-squares solutions that give an indication of the validity of the underlying admixture model (*i.e.,* do present-day Indian zebu breeds and the Guinean N'Dama population serve as adequate surrogates for the original parental populations?).

Hybrid composition of individual animals: The population genetic resolution that can be achieved using microsatellites has been demonstrated with previous studies where phylogenetic analysis was carried out using individuals as operational taxonomic units (OTUs) (BOWCOCK *et al.* 1994; ESTOUP *et al.* 1995). It is therefore tempting to estimate the hybrid composition of individual animals using the presence or absence of zebu-diagnostic alleles. This approach may prove particularly valuable for gaining an insight into the genetic erosion of the trypanotolerant taurine populations in West Africa. Seven populations were analyzed: Guinean N'Dama, Guinea Bissau N'Dama, Senegalese N'Dama, Gambian N'Dama, Senegalese Gobra zebu, Mauritanian Maure zebu and Sudanese Kenana zebu. Each animal can be analyzed at each locus that possesses zebu-specific alleles, scoring a 1 for each chromosome where a diagnostic allele is present. With 10 zebu-diagnostic loci, this gives an individual zebu index ranging from zero to 20.

RESULTS

Allelic diversities, heterozygosities, gene diversities, Hardy-Weinberg equilibrium and genetic differentiation: In total, 168 alleles were detected across the 20 loci when they were screened in the 20 taurine and zebu cattle populations. This gives a mean number of alleles at each locus of 8.4. The total number of detectable alleles increased dramatically to 200 when the other bovid samples were screened (American bison, European bison and banteng). When the allelic data is compared among cattle populations, some interesting patterns emerge (Table 3). The mean number of alleles (MNA) observed over a range of loci for different populations is considered to be a reasonable indicator of genetic variation with the provisos that the popula-

TABLE 3

Observed heterozygosities, gene diversities and mean number of alleles for 20 populations

Breed	Geographical origin	n	H_o	H_e	MNA (full sample)[a]	MNA (uniform sample)[b]
Aberdeen Angus	Scotland	33	0.477 ± 0.019	0.470 ± 0.054	3.7	3.5
Hereford	England	34	0.465 ± 0.019	0.456 ± 0.064	3.8	3.6
Jersey	Channel Islands	34	0.440 ± 0.019	0.432 ± 0.059	3.4	3.1
Kerry	Ireland	40	0.479 ± 0.018	0.473 ± 0.044	3.4	3.1
British Isles taurine	Four breeds	141	0.466 ± 0.009	0.514 ± 0.057	5.2	—
Charolais	France	36	0.525 ± 0.019	0.549 ± 0.055	4.5	4.0
Friesian	Netherlands	40	0.551 ± 0.018	0.545 ± 0.052	4.8	4.3
Simmental	Switzerland	36	0.499 ± 0.019	0.480 ± 0.053	4.4	3.9
Continental Europe taurine	Three breeds	112	0.526 ± 0.011	0.552 ± 0.054	6.0	—
N'Dama	Gambia	58	0.584 ± 0.014	0.583 ± 0.049	5.6	4.7
N'Dama	Guinea	63	0.513 ± 0.014	0.503 ± 0.053	4.3	3.8
N'Dama	Guinea Bissau	54	0.538 ± 0.015	0.534 ± 0.054	4.7	3.9
N'Dama	Nigeria	19	0.529 ± 0.026	0.556 ± 0.050	4.8	4.8
N'Dama	Senegal	48	0.532 ± 0.016	0.550 ± 0.053	5.4	4.5
West African taurine	Five populations	242	0.540 ± 0.007	0.549 ± 0.051	6.4	—
Butana	North Sudan	24	0.652 ± 0.022	0.599 ± 0.036	4.3	4.1
Kenana	South Sudan	38	0.643 ± 0.017	0.619 ± 0.041	4.9	4.6
Gobra	Senegal	59	0.633 ± 0.014	0.634 ± 0.045	5.9	4.8
Maure	Mauritania	55	0.631 ± 0.015	0.658 ± 0.040	6.2	5.3
White Fulani	Nigeria	24	0.610 ± 0.022	0.636 ± 0.044	4.8	4.7
African zebu	Five breeds	200	0.634 ± 0.008	0.652 ± 0.039	7.0	—
Hariana	North India	10	0.550 ± 0.035	0.548 ± 0.048	4.2	—
Sahiwal	Pakistan	13	0.542 ± 0.031	0.569 ± 0.043	4.1	—
Tharparker	North India	10	0.635 ± 0.034	0.575 ± 0.046	3.6	—
Asian zebu	Three breeds	33	0.573 ± 0.019	0.588 ± 0.044	5.2	4.8
Total	20 breeds	728	0.551 ± 0.004	0.643 ± 0.045	8.4	—

Sample sizes (n) are the number of animals typed and pooled samples are underlined. Observed heterozygosities (H_o) and gene diversities (H_e) are shown with their respective standard errors.

[a] The mean number of alleles (MNA) was calculated using all the animals in each population sampled.

[b] Uniform population sizes were used for the second column of MNA values. A random sample of 20 individuals was used for each population except for the three Indian breeds and the Nigerian N'Dama. In the case of the Indian breeds, a pooled sample of 20 different animals from the three breeds was used. In the case of the Nigerian N'Dama, the total sample of 19 animals was used.

tions are at mutation-drift equilibrium and that sample size is more or less the same for each population (NEI 1987). When the present data were analyzed, a positive correlation was observed between the MNA and the sample size of each population (Pearson product-moment correlation = 0.547). Consequently, to remove any sample bias, the MNA for a random sample of 20 animals was calculated for each population except the Nigerian N'Dama and the three Indian zebu populations. In the case of the Nigerian N'Dama, the original sample of 19 animals was used, and in the case of the three Indian zebu breeds, a random pooled sample of 20 animals was used to represent a combined Indian zebu population. The marked pattern of allelic diversity based on geographic origin evident when all animals were typed for each population was maintained when only 19–20 animals were used for each population. Therefore, the marked pattern of allelic diversity based on geographical origin shown in Table 3 is more likely to reflect distinctive population histories rather than a sampling artifact.

Among the different biogeographical groupings the four breeds from the British Isles exhibited the greatest allelic paucity and the African zebu breeds displayed the highest allelic diversity. The total pooled sample (n = 200) for the five African zebu breeds possessed a significantly higher MNA than the pooled sample (n = 241) from seven European breeds (Wilcoxon's signed rank test, P = 0.020). Two breeds showed the lowest number of detectable alleles with a mean of 3.4 per locus (3.1 when n = 20). These were the Kerry breed from southwest Ireland and the Jersey breed from the island of Jersey in the Channel Islands. This reduction in genetic variation may be attributable to genetic isolation, historical population bottlenecks or founder effects. Kerry cattle are an endangered indigenous breed that experienced a severe population bottleneck within the last 20 years (CURRAN 1990). The Jersey cattle

breed originate from a small island and have existed in effective genetic isolation because of strict breeding practices for ~200 years (FRENCH *et al.* 1966). The population with greatest allelic diversity was the Maure breed from Senegal in West Africa with a MNA of 6.2 (5.3 when $n = 20$). This population, in common with all African zebu breeds, has been influenced by historical zebu-taurine crossbreeding and the high allelic diversity observed is undoubtedly an artifact of admixture and the consequent input of both taurine and zebu alleles.

Observed heterozygosities and gene diversities: Table 3 shows mean heterozygosities computed across the 20 loci for each breed and biogeographical grouping. The mean observed heterozygosities ranged from 0.440 ± 0.019 for the Jersey breed to 0.652 ± 0.022 for the Butana breed. The overall pattern is similar to the mean number of alleles; the populations from the British Isles displaying the lowest observed heterozygosities and the African zebu breeds displaying the highest. The pooled African zebu sample from five breeds possess a significantly higher observed heterozygosity than the pooled European sample consisting of seven breeds (Wilcoxon's signed rank test, $P = 0.002$). The mean estimated gene diversities ranged from 0.432 ± 0.059, again for the Jersey breed, to 0.658 ± 0.040 for the Maure breed. The overall pattern concords with the previous observations that, for the most part, cattle breeds from the British Isles show the least genetic variation and crossbred African zebu populations display the most genetic variation. Again, the pooled European sample displayed a significantly lower value for this measure of diversity than the pooled African zebu sample (Wilcoxon's signed rank test, $P = 0.024$).

Perturbations from Hardy-Weinberg equilibrium: Figure 2 shows a diagrammatic summary matrix of the tests performed for deviations from HWE. Three tests were performed for each locus/population combination. In total, 33 locus/population comparisons gave one or more significant result for each test for HWE. Of these, 15 were instances when all three tests concurred at the $P \leq 0.05$ level, seven were instances when two tests concurred at $P \leq 0.05$ level and 11 were cases where only one test gave a significant result. Out of a total of 1200 tests, 70 gave a significant result, which is not many more than the 60 that would be expected by chance. However, when the results were pooled across loci and breeds using Fisher's combined probability method, a small number of consistent deviations emerged. The Aberdeen Angus and the Maure breeds gave a significant result at the $P \leq 0.01$ level for all three pooled tests. It is difficult to envisage factors that have caused such significant genetic substructure in the Aberdeen Angus breed. However, the Maure breed was collected as two separate samples in different years (24 animals in 1992 and 31 animals in 1993). In addition, these cattle were sampled from a dwindling population in Senegal and not in the breed's country of origin (Table 1). These animals may have been representatives of a number of small insular clusters. Either of these factors alone would tend to increase the likelihood of deviations from HWE and taken together, they may account for the observed disequilibrium.

An exact test for population differentiation: When the test for population differentiation was performed for the 190 pairwise population comparisons, the results were essentially uninformative. In all but one case, the probabilities were highly significant ($P \ll 0.01$) that the two populations being compared were differentiated. The only two populations that showed a remote genetic affinity using this test were the N'Dama sampled in Nigeria and the N'Dama collected in Gambia ($P = 0.036$). It is known that the Nigerian population was originally derived from Gambian imports brought to Nigeria within the last 20 years. This recent common identity may explain the relatively high P value observed for the comparison between these two populations.

An evolutionary tree for domesticated cattle constructed from microsatellite data: A neighbor-joining phylogeny was constructed using a matrix of D_{SW} genetic distances among the 20 cattle populations and the three related species. However, the topology of this tree was heavily distorted by the inclusion of populations known to be admixtures between zebu and taurine populations. Because hybrid populations are not discrete evolutionary lineages, trees with admixed populations violate the fundamental principles of phylogeny reconstruction (FELSENSTEIN 1982; NEI 1987).

Figure 3 shows a neighbor-joining tree subsequently constructed from a matrix of D_{SW} distances using only 12 non-admixed cattle populations and the three related species. All of the African zebu breeds were excluded for the purposes of this tree. Also excluded were three of the N'Dama populations known to be crossbred from previous *Y*-chromosome studies and the admixture data presented in a later section. Figure 3 clearly illustrates the extent of the divergence between the taurine and zebu clades. A 97% bootstrap probability supports the separation of the *Bos taurus* and *Bos indicus* lineages. The separation between the European and African taurine groups is supported by a 95% bootstrap value, indicating that these two groups are also quite distinct evolutionary lineages. The internal structure of the European clade is not strongly supported by the bootstrap tests. This probably reflects a shallow depth for the radiation of European cattle and historical admixture among the breeds. Three of the microsatellites screened are known to be tightly physically linked on chromosome *23* (BoLA DR2B, BoLA DRP1 and PRL, EGGEN and FRIES 1995). Therefore an additional phylogenetic tree was constructed using the other 17 markers plus the most polymorphic of the three linked microsatellites (BoLA DRP1). The topology of this new

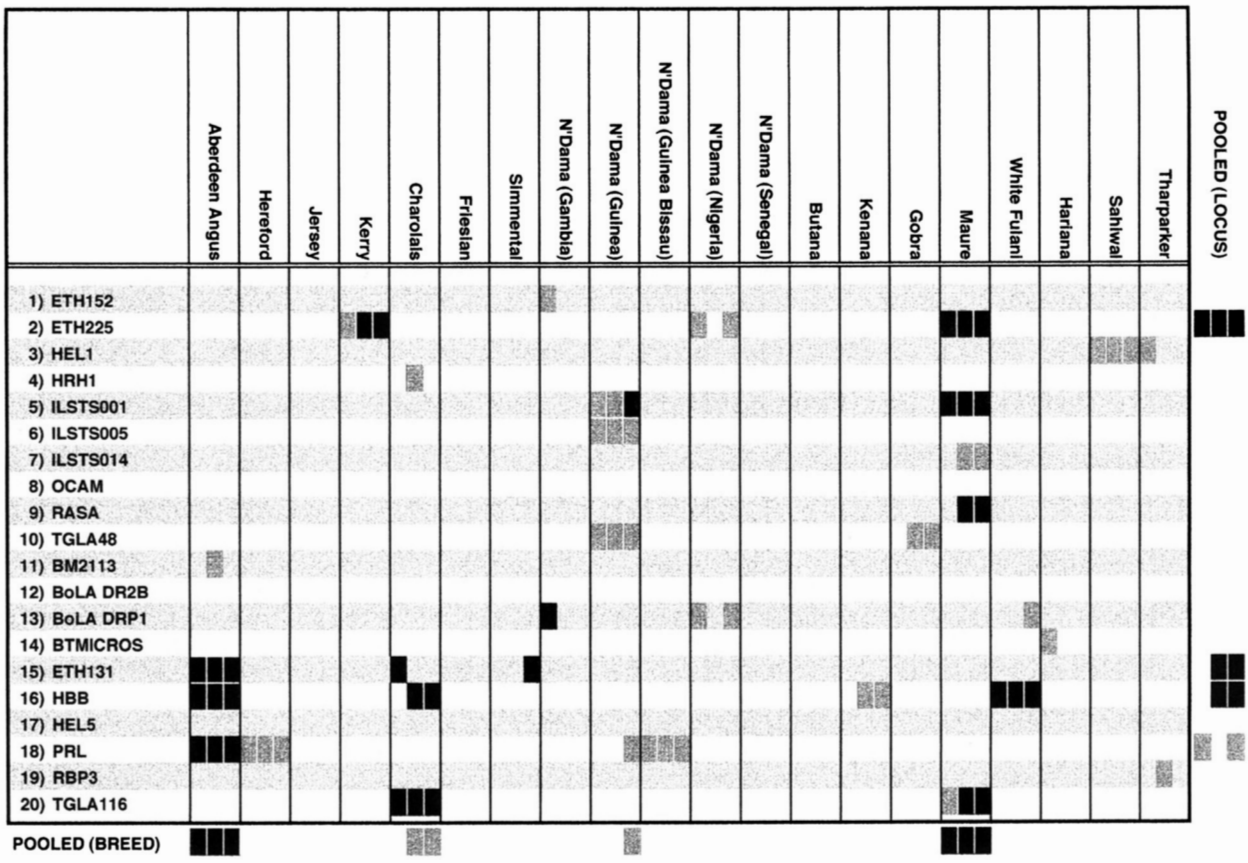

FIGURE 2.—Summary matrix of three tests performed for HWE proportions for all 400 locus/population combinations. The results of the shuffled chi-square test are presented in the first cell of each comparison, the shuffled likelihood ratio (L) test results are shown in the second cell of each comparison and the Markov chain exact probability test results are shown in the third cell. A dark gray cell indicates a significant deviation at the $P \leq 0.05$ level and a solid black cell indicates a significant result at the $P \leq 0.01$ level. Pooled results across loci and across breeds were calculated using Fisher's method for combining probabilities (see MATERIALS AND METHODS).

tree was identical to that shown in Figure 3 and the bootstrap values were very similar (data not shown).

Estimation of times of divergence: It has been shown that D_{SW} is approximately linear with respect to time (SHRIVER *et al.* 1995; L. JIN, unpublished data). Using the two bison species as an outgroup, it is therefore possible to obtain approximate estimates of the time of divergence between *Bos taurus* and *Bos indicus* and also between the two major types of taurine cattle. The estimated time since divergence between Bison and Bos is 1.4 million years based on allozyme data (HARTL *et al.* 1988), or at least one million years based on paleontological evidence (A. W. GENTRY, personal communication). Using this information and all the pairwise genetic distance values among the bison, taurine and zebu clades, a divergence time between *Bos taurus* and *Bos indicus* was extrapolated as between 610,000 to 850,000 years, depending on which estimate for the Bison/Bos split was employed. In a similar manner, the divergence between the African and European taurine clade was estimated as between 180,000 to 250,000 years.

Principal component analysis: Figure 4 illustrates the first, second and third principal components (PC) for the microsatellite allele frequency distributions in 20 cattle populations. The first PC accounts for 52.4% of the variation and clearly distinguishes the taurine and zebu groups. The second PC, which also differentiates the taurine and zebu groups, summarizes 14.4% of the variation. The third PC describes 7.2% of the variation and clearly separates the European and African taurine populations. In the three-dimensional space encompassed by the first three PCs, the African zebu populations form an almost linear array stretching towards the Asian zebu cluster at one end and to the five N'Dama populations at the other end. On close inspection, the position of the various African populations corresponds to the relative proportion of zebu and taurine alleles in their gene pools (see next section). This distinctive pattern provides a clearer representation of the true evolutionary history of the admixed zebu populations than a phylogenetic reconstruction.

Zebu genetic introgression and admixture in African cattle populations: Table 4 shows the percentage of zebu-specific alleles at 10 diagnostic loci in five African zebu breeds, five African taurine populations and the pooled Asian zebu sample. Also shown are the zebu

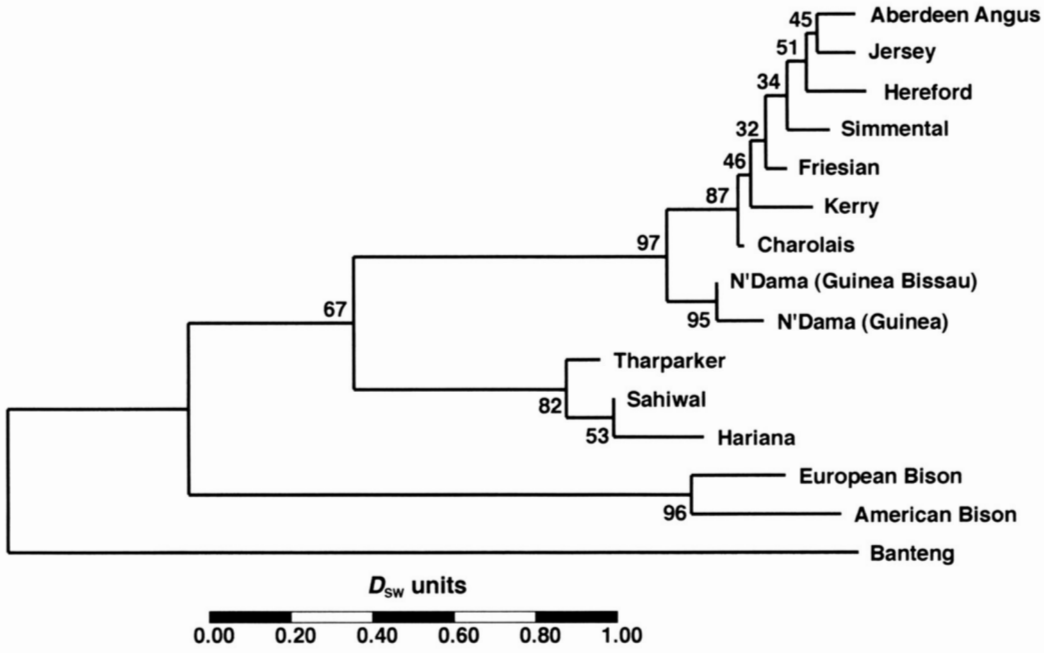

FIGURE 3.—Neighbor-joining tree summarizing D_{SW} genetic distances among 12 non-admixed cattle populations and three related species. Bootstrap values indicating the degree of support for each branch point are shown beside the node as the percentage of replicates in which the cluser to the right of the node was recovered. The linear scale relates the branch lengths to units of D_{SW}. The root of the tree was placed at the midpoint of the longest branch separating the banteng from the other groups.

admixture proportions calculated using the gene identity method. It is interesting to note that the distribution of zebu alleles and the zebu admixture proportions decline from East to West Africa and then follow a steep north-south gradient in West Africa. The five N'Dama populations display varying degrees of zebu influence depending on their proximity to the edge of the subtropical Guinean biome, the area with a high tsetse challenge. This pattern of hybridization and genetic exchange reflects the original historical influx of Asian zebu cattle into East Africa and also the lack of zebu penetration into areas where tsetse-borne trypanosomiasis is endemic (EPSTEIN 1971). The correlation coefficients calculated from the weighted least square solutions for the admixture proportions are all highly significant, indicating that the three North Indian zebu breeds and the Guinean N'Dama population are appropriate surrogates for the original source populations that contributed to the genetic exchange between taurine and zebu cattle in Africa.

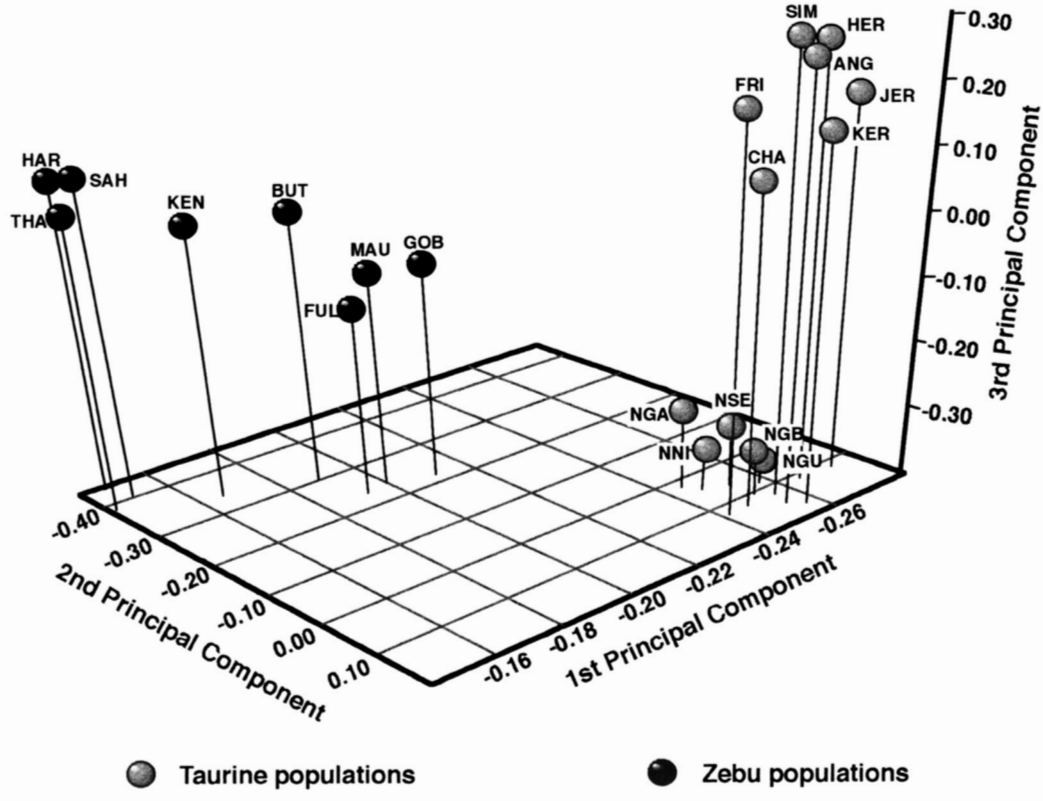

FIGURE 4.—Principal components diagram showing the first three PCs from transformed allele frequencies in 20 cattle populations. The three-letter code corresponds to the populations sampled as follows: ANG, Aberdeen Angus; HER, Hereford; JER, Jersey; KER, Kerry; CHA, Charolais; FRI, Friesian; SIM, Simmental; NGA, N'Dama (Gambia); NGU, N'Dama (Guinea); NGB, N'Dama (Guinea Bissau); NNI, N'Dama (Nigeria); NSE, N'Dama (Senegal); BUT, Butana; KEN, Kenana; GOB, Gobra; MAU, Maure; FUL, White Fulani; HAR, Hariana; SAH, Sahiwal; THA, Tharparker.

TABLE 4

Zebu-diagnostic alleles and zebu admixture proportions in African and Asian populations

Breed	Geographical origin	Zebu alleles[a] (%)	Admixture proportions (%)[b]
Asian zebu	Three breeds	67.0 ± 3.6	100[c]
Kenana	South Sudan	55.8 ± 3.5	83.2 ± 2.6 (0.988)
Butana	North Sudan	47.3 ± 4.5	73.6 ± 1.7 (0.925)
White Fulani	Nigeria	43.3 ± 4.0	65.4 ± 1.8 (0.953)
Maure	Mauritania	39.7 ± 2.9	61.3 ± 1.8 (0.962)
Gobra	Senegal	36.4 ± 2.7	59.8 ± 1.3 (0.981)
N'Dama	Nigeria	8.7 ± 2.8	14.8 ± 1.5 (0.994)
N'Dama	Gambia	8.6 ± 1.6	18.9 ± 1.6 (0.992)
N'Dama	Senegal	5.1 ± 1.4	12.4 ± 1.4 (0.995)
N'Dama	Guinea Bissau	1.5 ± 0.7	6.9 ± 2.0 (0.993)
N'Dama	Guinea	0.2 ± 0.2	0[c]

[a] The mean percentage of zebu-specific alleles across 10 diagnostic loci. Standard errors were computed as the standard error of a binomial distribution with sample size equal to the number of chromosomes tested (10 diagnostic loci, therefore: 20 chromosomes × population size).

[b] Admixture proportions calculated from all allele frequencies using the gene identity method with standard errors calculated according to CHAKRABORTY (1985). Weighted least-squares correlation coefficients are shown in parentheses.

[c] Source population.

A high-resolution map of zebu introgression in West African cattle: Figure 5 shows a map that links the approximate geographical location of each of 337 animals sampled in and around the Senegambia region of West Africa to the proportion of zebu alleles present at 10 zebu-diagnostic microsatellites in each individual genome. Also shown for comparative purposes are the 38 Kenana animals sampled in southern Sudan. The north-south gradient of zebu introgression running through the semi-desert Sudano-sahelian biome into the subtropical Guinean region is clearly evident. The N'Dama population sampled at the edge of the tsetse zone in Gambia have an average of 8.6% alleles that are zebu in origin. Only 12.1% (7/58) of these animals did not display any zebu alleles. This contrasts with a N'Dama population sampled 350 km further south, deep within the equatorial tsetse zone. These Guinean N'Dama have an average of 0.2% zebu-specific alleles with 96.8% (61/63) of the cattle showing no trace whatsoever of zebu influence. Intermediate values were noted for the two populations situated between these two extremes. The Senegalese sample displayed 5.1% zebu-specific alleles with 41.7% (20/48) of the sample giving no indication of zebu ancestry. Cattle sampled in Guinea Bissau showed 1.5% zebu-specific alleles and 72.2% (32/54) of these animals did not betray any zebu influence.

DISCUSSION

Domesticated cattle have been the subject of numerous surveys of genetic variation within and among populations (for review see BAKER and MANWELL 1991). Previous studies have tended to focus on European breeds of cattle using allozyme or immunoprotein techniques. The first DNA-based surveys did not appear until the 1990's and these studies were generally broader in scope, encompassing both European and non-European populations (SUZUKI *et al.* 1993; BRADLEY *et al.* 1994, 1996; LOFTUS *et al.* 1994a,b). These reports have been based on analysis of mtDNA or *Y*-chromosome haplotypes and the present study is the first comprehensive survey of genetic variation at autosomal nuclear DNA markers in domesticated cattle.

Microsatellite variation and cattle genetic diversity: Striking patterns were evident in the distribution and pattern of genetic variation observed among the 20 cattle populations examined.

The seven European cattle breeds and particularly those from the British Isles displayed reduced levels of genetic variation as estimated by allelic diversity, observed heterozygosity and gene diversity. These observations may be a consequence of the recent evolutionary history of European cattle populations. Genetic isolation and biological manipulation by humans have been consistent features of the development of European breeds since the original demarcation of landraces into recognized breeds ~150 years ago (FELIUS 1995). The introduction of scientific breeding and reproductive technology in the 1950's has accelerated these processes, and in many respects, European cattle populations have a unique genetic history that may have given rise to the attenuated patterns of genetic variation observed in extant breeds.

Half (10) of the microsatellite markers displayed markedly different allelic distributions when analyzed in taurine and zebu cattle populations. These 10 loci showed a single allele or group of alleles that were present at high frequencies in Asian zebu breeds, intermediate frequencies in crossbred African zebu populations, low frequencies in African taurine populations

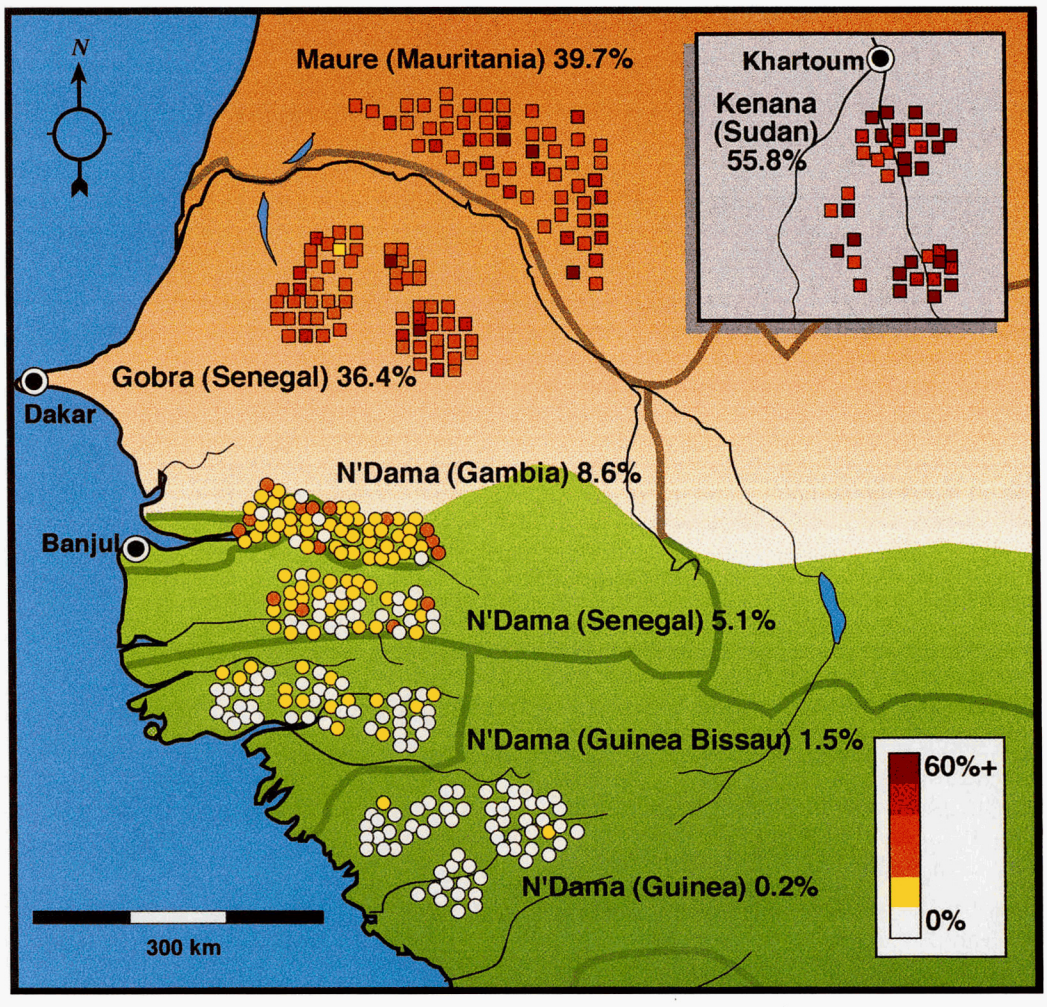

FIGURE 5.—Genetic introgression of zebu-specific alleles in West African cattle. Individual cattle are represented by circles (taurine animals) or squares (zebu animals). The color-coded scale represents the proportion of zebu-specific alleles in percentage terms from an original scale of zero to 20. The percentage beside each population is the proportion of zebu-specific alleles in the population as a whole (Table 4). With the exception of the Maure, the position of each animal corresponds approximately to the location where they were sampled. Although the Maure breed was actually sampled in Senegal, they are recent migrants to this country and were relocated to Mauritania for the purposes of this diagram. For comparative purposes, the gray window shows the East African zebu Kenana breed that was sampled in Sudan. The green shaded area denotes the arboreal tsetse-infested zone.

and were either absent or present at very low frequencies in European taurine populations (Figure 1). The presence of these zebu-specific or diagnostic alleles underlines the substantial divergence between the taurine and zebu genomes and supports the conclusions of LOFTUS *et al.* (1994a,b) concerning the separate origins of domesticated taurine and zebu cattle.

Another aspect of the biological history of Indian zebu cattle revealed by the distribution of these diagnostic alleles is the suggestion that the zebu breeds sampled in northern India may have been subject to limited genetic exchange and admixture with taurine cattle. This hypothesis emerges because taurine-specific alleles equivalent to those present in zebu genomes were not detected in this study. If the two groups had followed mutually exclusive evolutionary pathways since common ancestry, symmetrical divergence would be expected and European taurine should display high frequency alleles that are not present in pure Indian zebu genomes. It is likely that these alleles do actually exist, but that taurine introgression has introduced them into northern Indian cattle populations and masked their exclusivity in this survey. In effect, the Indian subcontinent may represent a mirror-image of the situation in Africa. Zebu populations sampled in the north of the region may have been influenced by early crossbreeding due to trade and migration between the Fertile Crescent and the Indian subcontinent (KULKE and ROTHERMUND 1990). On the other hand, zebu populations originating further south would have been less exposed to taurine introgression from northern Asia and, at any particular zebu-diagnostic locus, they should display zebu-specific allele frequencies that are closer to unity or fixation. Ongoing analysis of microsatellite variation in zebu breeds from southern India supports this scenario and suggests that a cline of taurine introgression may run from the north to the south of the Indian subcontinent (C. GALLIARD, personal communication).

Implications for the evolutionary origins of domesticated cattle: The most widespread view of the origins of domesticated cattle is that the *Bos taurus* and *Bos indicus* subspecies originate from the same early Neolithic domestication centers (EPSTEIN 1971; EPSTEIN and MASON 1984; PAYNE 1991). The pervasive influence of this viewpoint is evidenced by the classification of zebu cattle as *Bos taurus* in a standard text on mammalian taxa (NOWAK 1991).

Phylogenetic analysis of microsatellite polymorphisms in domesticated cattle has underlined the substantial divergence between zebu and taurine breeds. Previous studies of allozyme and immunoprotein variation have also indicated a large divergence between the two cattle subspecies (BRÆND 1972; NAIK 1978; MANWELL and BAKER 1980; VON GRAML *et al.* 1986). However, until mtDNA sequence data became available, no attempt was made to date the time of the original split. Analysis of mtDNA control region sequence variation using bison sequence as an outgroup suggested that the two subspecies diverged a minimum of 210,000 years ago (LOFTUS *et al.* 1994a; BRADLEY *et al.* 1996). This work may be criticized on the basis that phylogenies derived from a single genetic locus may not reflect organismal evolution (PAMILO and NEI 1988; BALL *et al.* 1990; WU 1991). However, the total array of DNA-based information supporting a deep split between the taurine and zebu clades now includes mtDNA sequence or RFLP haplotypes from 130 cattle (LOFTUS *et al.* 1994b; BRADLEY *et al.* 1996), *Y*-chromosome haplotypes in 184 cattle (BRADLEY *et al.* 1994), and the allele frequency variation at 20 independent autosomal microsatellites in 728 cattle reported here. A substantial body of archaeological evidence also exists that provides support for a separate origin of zebu cattle somewhere on the Indian subcontinent (ALLCHIN 1969; GRIGSON 1980; MEADOW 1984; MEADOW 1993). The alternative hypothesis supporting a recent evolution of zebu cattle from domesticated taurine ancestral populations now seems untenable. One problem remains however, the discrepancy between the microsatellite-derived divergence time and the date estimated from mtDNA sequence evolution (BRADLEY *et al.* 1996). The microsatellite estimate (610,000–850,000 years ago) is approximately three times larger than the upper bound of the date derived from mtDNA sequence analysis (275,000). This disparity may reflect differences in evolutionary dynamics or, more plausibly, may result from the potentially large errors inherent in these estimates.

Recent mtDNA analysis has tentatively suggested that African and European taurine cattle may have also had different domestic origins (BRADLEY *et al.* 1996). Analysis of microsatellite variation in European and African taurine cattle also provides some support for this interpretation. In particular, the tree constructed from D_{SW} genetic distances indicates that the African and European taurine gene pools diverged between 150,000 and 250,000 years ago. Resolution of the microsatellite variation into principal components also emphasizes a clear separation between the African and European taurine groups.

A number of workers have presented archaeological evidence for the presence of domesticated cattle in North Africa during the early stages of the Neolithic period (CARTER and CLARK 1976; GAUTIER 1987; GRIGSON 1991; WENDORF and SCHILD 1994). Although the genetic data provide some support for an autochthonous domestication of African cattle, genetic exchange between populations emerging from the Middle East and indigenous African aurochs cannot be discounted. Until unambiguous evidence for the domestication process such as faunal shifts or clear size diminution is discovered, it is unlikely that this issue will be resolved through standard archaeological methods. However, molecular studies, in the form of analysis of ancient DNA from faunal remains may provide new insights into this problem.

Male-mediated zebu genetic introgression in African cattle populations: The unique history of African domesticated cattle is reflected in both the morphology and genetic composition of extant populations. For the most part, the original taurine substrate of African cattle has been displaced and absorbed by successive waves of zebu migration from Asia and Arabia. This historical influx is evident in the cline of zebu genomic identity observed running in an east-west direction across Africa and in a north-south direction in West Africa (Table 4).

The genetic composition of African cattle populations has been assayed using three different genomic systems. Mitochondrial studies of European, African and Asian cattle have revealed two distinct mtDNA lineages, one taurine and one zebu. These lineages were detected using both control region sequence differences and RFLP analysis of the whole mtDNA molecule (SUZUKI *et al.* 1993; LOFTUS *et al.* 1994a,b; BRADLEY *et al.* 1996). In total, mtDNA haplotypes (mitotypes) have been characterized from 90 African animals (35 taurine and 55 zebu) originating from both West and East Africa. Without exception, all of these animals possessed a taurine mitotype. In contrast to this, analysis of the *Y* chromosome in African populations using a bovine *Y*-specific DNA probe (btDYZ-1) and a *Y*-specific RAPD marker (ILO65) has shown a markedly different pattern (BRADLEY *et al.* 1994; TEALE *et al.* 1995). A zebu *Y*-chromosome haplotype has become widespread throughout Africa and has reached high frequencies in breeds and populations nominally classified as pure taurine. These asymmetrical distributions of the uniparental inherited genetic systems, highlight the male-mediated nature of zebu gene flow on the African continent. The autosomal genomic composition of African cattle populations presents an intermediate picture between the two uniparental extremes and represents the overall penetration of the zebu genome most accurately. Figure 6 shows a summary diagram of zebu genetic introgression for the three systems.

Although they lack zebu mtDNA, as shown in Table 4 and Figure 6, the two East African zebu populations that were surveyed for microsatellite variation (Butana and Kenana) have predominantly zebu autosomal genomes with admixture proportions of 73.6 ± 1.7% and 83.2±2.6%, respectively. They also display zebu *Y* chro-

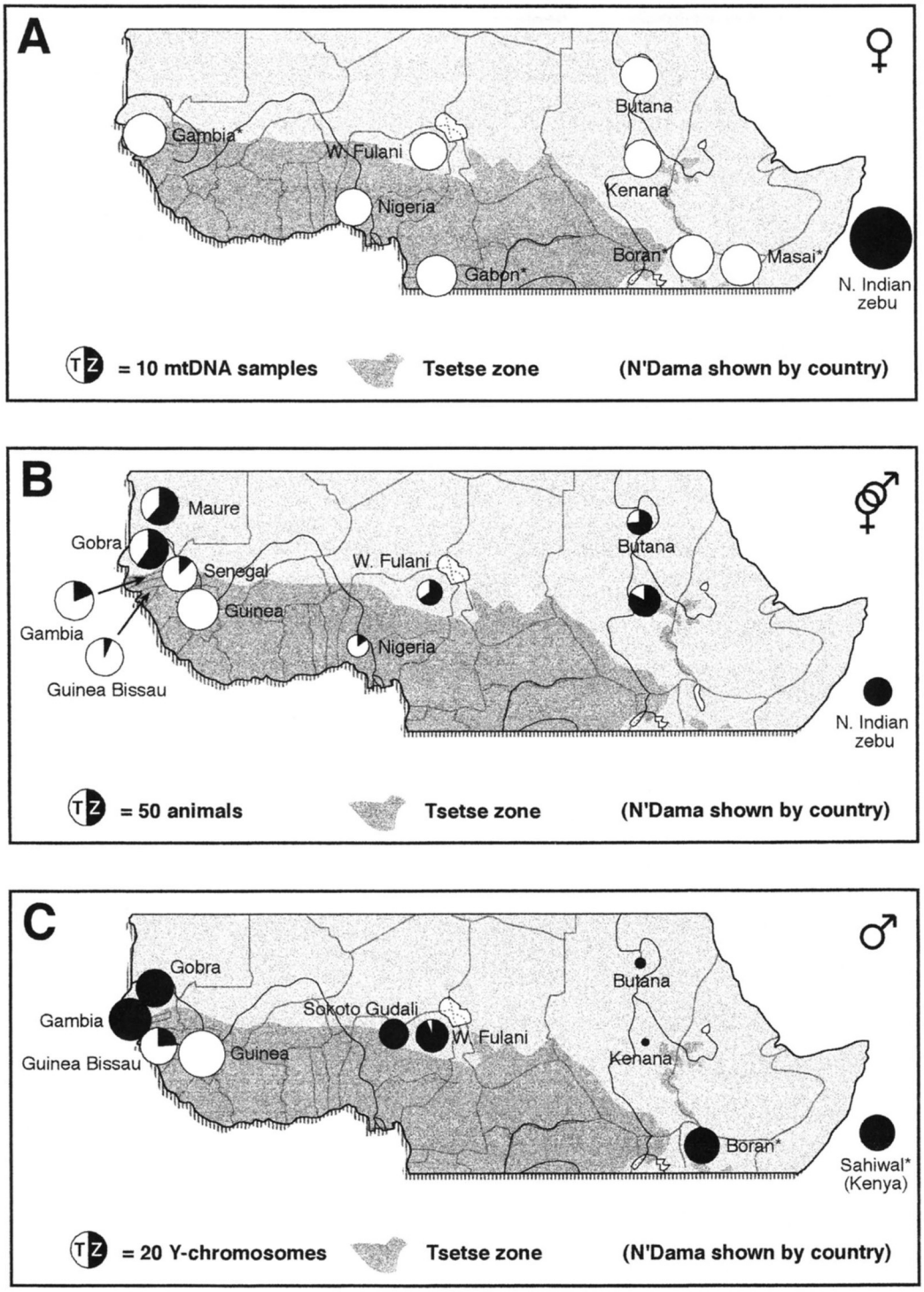

FIGURE 6.—The introgression of three different zebu genomic components into African cattle. Sample size is proportional to the area of each pie-circle as indicated on each of the three diagrams. The distribution of taurine and zebu mitotypes are shown in A. Zebu admixture proportions are shown in B. The black portion of each circle represents zebu admixture proportions from Table 4 and the area of each circle is proportional to sample size. The introgression of the zebu *Y* chromosome is shown in C. Again, the black portion of the circles corresponds to the proportion of zebu *Y* chromosomes in the population. Data for A and C were taken from SUZUKI *et al.* (1993), BRADLEY *et al.* (1994), LOFTUS *et al.* (1994b) and TEALE *et al.* (1995). Populations marked with an asterisk (*) were not analyzed in our laboratory. Sample sizes are not given in TEALE *et al.* (1995) and for the purposes of C an arbitrary size of 20 is assumed for the Boran and Kenyan Sahiwal populations.

mosomes. The West African zebu breeds show a similar pattern, although with less zebu influence on their autosomal genomes. The most striking result is that obtained for the Gambian N'Dama. This population is morphologically taurine and was sampled on the fringes of the tsetse zone near the Gambia river. Their autosomal genome is composed of 8.6 ± 1.6% zebu alleles with a zebu admixture proportion of almost 20% (18.9 ± 1.6%). In common with all the other African populations, they possess no zebu mitotypes. However, all 28 males screened with the btDYZ-1 *Y*-specific probe yielded zebu haplotypes (Figure 6 and BRADLEY *et al.* 1994). This result dramatically illustrates the asymmetrical nature of zebu gene flow, highlighting the genetic inertia of the African taurine mtDNA gene pool and the rapid dispersal of the zebu *Y* chromosome. It suggests a mechanistic pattern for zebu genetic introgression that probably applies to the whole of Africa.

There are number of factors that could account for the unusual dynamics of zebu gene flow in African cat-

tle. However, it seems clear that the genetic exchange between zebu bulls and taurine females may have been particularly important. The unusual population structure of domesticated cattle with consequent disproportionate contributions to future generations from individual males may have also contributed to the observed pattern. The Arab peoples who brought zebu cattle to Africa commencing about 700 AD may, for economic or logistical reasons, have brought disproportionate numbers of males. Zebu cattle would have been prized for their ability to withstand arid conditions and bulls in particular may have been particularly valuable. A single zebu bull could have been crossed with a herd of taurine females to produce an arid-resistant hybrid F_1 population that possessed no zebu mitotypes and males fixed for the zebu *Y* chromosome. Under these circumstances, the composition of the African gene pool may have been transformed very rapidly and nuclear genomes could have become progressively more zebu without any appreciable input of zebu mitotypes.

Another factor that may have hastened the spread of the zebu nuclear genome was the periodic rinderpest epizootics that were introduced from Asia with zebu cattle (EPSTEIN 1971). Through longer exposure to the disease, zebu cattle are more resistant to the effects of rinderpest than taurine cattle (PAYNE 1970). The most destructive rinderpest outbreak to happen within the last 200 years was that of 1890–91. During this period, over 80% of the livestock of East Africa were destroyed and many wild species were also decimated (EPSTEIN 1971). The majority of the cattle left after this disease episode had zebu ancestry and the last vestiges of pure East African taurine cattle were practically eradicated. During the time since the large-scale influx of zebu cattle about 1300 years ago, there would have been a number of rinderpest outbreaks and it is possible to envisage a precipitous decline in taurine population numbers during each epizootic.

Implications for genetic conservation and breed characterization: The introgression of the zebu genome into trypanotolerant populations of taurine N'Dama cattle shown in Figure 5 illustrates the precarious state of some of these populations. The sharp cline in zebu-specific alleles at the tsetse zone transition reflects the disease challenge faced by zebu animals inhabiting the equatorial region. It also suggests that the introduction of zebu alleles into the southern forested areas must be predominantly through admixed taurine animals. The effects of deforestation, desertification and the subsequent migration of zebu populations are apparent. If tsetse densities decrease because of destruction of the species' natural habitat, even populations such as the Guinean N'Dama would quickly succumb to genetic absorption and introgression from zebu populations moving into deforested areas. The mountainous Fouta Djallon region in Guinea is reputed to be the area where the N'Dama breed originate (FELIUS 1995). The Guinean population surveyed were taken from these mountains and may represent the last remnants of pure N'Dama left in West Africa. It is therefore imperative that this population is targeted in future conservation efforts.

The systematic classification of cattle genetic resources in Africa has been actively researched for a long period of time and a broad consensus has yet to be reached (MASON 1951; MASON and MAULE 1960; PAYNE 1970; EPSTEIN 1971; BAKER and MANWELL 1980; MAULE 1990; GRIGSON 1991). Previous commentators have been preoccupied with traits of dubious systematic value such as horn shape, skull morphology and the location of the hump in zebu/taurine crossbreds. The molecular data demonstrate that many of the different physical characteristics of African cattle (for example, thoracic *vs.* cervico-thoracic humps) are probably just manifestations of the variable penetrance of admixed zebu and taurine morphological traits.

Based on analysis of mtDNA, no African zebu breed should be classified as true zebu. The retention of taurine mitotypes is *de facto* evidence for some input of taurine genes during the development of these breeds. It may actually be more appropriate to adopt the term "zeboid," first suggested by MASON (1951). Likewise, many taurine populations should not be considered purebred based on the infiltration of zebu *Y* chromosomes and autosomal zebu-specific alleles. Accurate estimates of admixture in African cattle would therefore seem to provide the most rational biological framework for classification of African cattle populations. A set of zebu-diagnostic microsatellites could be used to disentangle and simplify classification systems. The proportion of zebu admixture could be correlated with production and fitness-related characteristics such as fertility, feed-intake and conversion, growth rate, milk production, water requirements, disease resistance and mortality. The relative ease with which microsatellites may be screened in a large number of samples makes this approach particularly attractive.

Conclusions: This study demonstrates the resolution that can be achieved using microsatellite polymorphisms for studies of genetic microdifferentiation. The primary outcome was the affirmation of the large divergence between the *Bos taurus* and *Bos indicus* lineages previously reported for allozyme and mitochondrial data (MANWELL and BAKER 1980; LOFTUS *et al.* 1994a,b; BRADLEY *et al.* 1996). This provides strong support for the dual-domestication theory proposed by LOFTUS *et al.* (1994a). The discordance observed among three different genetic systems in African cattle highlights the potential for misinterpretation if mtDNA is used as the sole source of data when investigating phylogeographic processes. Important sex-related demographic patterns may be overlooked unless a broad approach, encompassing mtDNA, *Y*-chromosome polymorphisms and autosomal variation is used.

We acknowledge the assistance of the following people with sample collection and data analysis: A. M. BADI, D. S. BALAIN, M. DIALLO, H. DJATA, A. FAYE, B. FYE, A. GUEYE, A. HASSAN, L. JIN, A. T. LLOYD, C. MEGHEN, M. MURRAY, L. O. NGERE, J. NORRIS, T. OTA, J. RIZGALLA, V. J. SHANKER, P. M. SHARP, B. SAUVEROCHE, R. S. SOW, M. J. STEAR, C. S. TROY, H. WAGNER, T. WILSON, Y. ZHONG. This work was funded by the European Commission DGX11 Science and Technology in Developing Countries Program. D.E.M. is supported by a Research Fellowship in Bioarchaeology from the Wellcome Trust. M.D.S. is supported by a Fellowship from the W. M. Keck Foundation for Advanced Training in Computational Biology and a grant from the National Institute of Justice.

LITERATURE CITED

ALLCHIN, F. R., 1969 Early domestic animals in India and Pakistan, pp. 317–322 in *The Domestication and Exploitation of Plants and Animals,* edited by P. J. UCKO and G. W. DIMBLEBY. Duckworth, London.

ANDERSSON, L., J. BÖHME, L. RASK and P. A. PETERSON, 1986 Genomic hybridization of bovine class II major histocompatibility genes: 1. Extensive polymorphism of DQα and DQβ genes. Anim. Genet. **17:** 95–112.

BAKER, C. M. A., and C. MANWELL, 1980 Chemical classification of cattle. 1. Breed groups. Anim. Blood. Grps. Biochem. Genet. **11:** 127–150.

BAKER, C. M. A., and C. MANWELL, 1991 Population genetics, molecular markers and gene conservation of bovine breeds, pp. 221–304 in *Cattle Genetic Resources,* edited by C. G. HICKMAN. Elsevier, Amsterdam.

BALL, R. M., J. E. NEIGEL and J. C. AVISE, 1990 Gene genealogies within the organismal pedigrees of random-mating populations. Evolution **44:** 360–370.

BOWCOCK, A. M., A. RUIZ-LINARES, J. TOMFOHRDE, E. MINCH, J. R. KIDD *et al.,* 1994 High resolution of human evolutionary trees with polymorphic microsatellites. Nature **368:** 455–457.

BRADLEY, D. G., D. E. MACHUGH, R. T. LOFTUS, R. S. SOW, C. H. HOSTE *et al.,* 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.

BRADLEY, D. G., D. E. MACHUGH, P. CUNNINGHAM and R. T. LOFTUS, 1996 Mitochondrial diversity and the origins of African and European cattle. Proc. Natl. Acad. Sci. USA **93:** 5131–5135.

BRÆND, M., 1972 Studies on the relationships between cattle breeds in Africa, Asia and Europe: evidence obtained by studies of blood groups and protein polymorphisms. World Rev. Anim. Prod. **8:** 9–14.

CARTER, P. L., and J. D. CLARK, 1976 Adrar Bous and African cattle, pp. 487–493 in Proceedings of 7th Pan-African Congress on Prehistory, Addis Ababa, 1971.

CAVALLI-SFORZA, L. L., P. MENOZZI and A. PIAZZA, 1994 *The History and Geography of Human Genes.* Princeton University Press, Princeton, NJ.

CHAKRABORTY, R., 1975 Estimation of race admixture—a new method. Am. J. Phys. Anthropol. **42:** 507–511.

CHAKRABORTY, R., 1985 Gene identity in racial hybrids and estimation of admixture rates, pp. 171–180 in *Genetic Microdifferentiation in Man and Other Animals,* edited by J. V. NEEL and Y. AHUJA. Indian Anthropological Assoc., Delhi, India.

CHAKRABORTY, R., 1986 Gene admixture in human populations: models and predictions. Yearbook Phys. Anthropology **29:** 1–43.

CHAKRABORTY, R., M. FORNAGE, R. GUEGUE and E. BOERWINKLE, 1991 Population genetics of hypervariable loci: analysis of PCR-based VNTR polymorphism within a population, pp. 127–143 in *DNA Fingerprinting: Approaches and Applications,* edited by T. BURKE, G. DOLF, A. J. JEFFREYS and R. WOLFF. Birkhäuser, Basel.

CHAKRABORTY, R., M. I. KAMBOH, M. NWANKWO and R. E. FERRELL, 1992 Caucasian genes in American Blacks: new data. Am. J. Hum. Genet. **50:** 145–155.

CURRAN, P. L., 1990 *Kerry and Dexter Cattle and Other Ancient Irish Breeds: A History.* The Royal Dublin Society, Dublin, Ireland.

DEKA, R., L. JIN, M. D. SHRIVER, L. M. YU, S. DECROO *et al.,* 1995 Population genetics of dinucleotide $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms in world populations. Am. J. Hum. Genet. **56:** 461–474.

EDWARDS, A., H. A. HAMMOND, L. JIN, C. T. CASKEY and R. CHAKRABORTY, 1992 Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics **12:** 241–253.

EGGEN, A., and R. FRIES, 1995 An integrated cytogenetic and meiotic map of the bovine genome. Anim. Genet. **26:** 215–236.

EPSTEIN, H., 1971 *The Origin of the Domestic Animals of Africa.* Africana, New York.

EPSTEIN, H., and I. L. MASON, 1984 Cattle, pp. 6–27 in *Evolution of Domesticated Animals,* edited by I. L. MASON. Longman, London.

ESTOUP, A., L. GARNERY, M. SOLIGNAC and J-M. CORNUET, 1995 Microsatellite variation in honey bee (*Apis mellifera L.*) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics **140:** 679–695.

FELIUS, M., 1995 *Cattle Breeds—An Encyclopedia.* Misset, Doetinchem, The Netherlands.

FELSENSTEIN, J., 1982 How can we infer geography and history from gene frequencies? J. Theor. Biol. **96:** 9–20.

FELSENSTEIN, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **29:** 783–791.

FRENCH, M. H., I. JOHANSSON, N. R. JOSHI and E. A. MCLAUGHLIN, 1966 *European Breeds of Cattle.* Food and Agricultural Organization of the United Nations, Rome.

GAUTIER, A., 1987 Prehistoric men and cattle in North Africa: a dearth of data and a surfeit of models, pp. 163–187 in *Prehistory of Arid North Africa: Essays in Honor of Fred Wendorf,* edited by A. E. CLOSE. Southern Methodist University Press, Dallas.

GOLDSTEIN, D. B., A. RUIZ LINARES, L. L. CAVALLI-SFORZA and M. W. FELDMAN, 1995a Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Natl. Acad. Sci. USA **92:** 6723–6727.

GOLDSTEIN, D. B., A. RUIZ LINARES, L. L. CAVALLI-SFORZA and M. W. FELDMAN, 1995b An evaluation of genetic distances for use with microsatellite loci. Genetics **139:** 463–471.

GRIGSON, C., 1980 The craniology and relationships of four species of *Bos.* 5. *Bos indicus* L. J. Arch. Sci. **7:** 3–32.

GRIGSON, C., 1991 An African origin for African cattle?—some archaeological evidence. Afric. Archaeol. Rev. **9:** 119–144.

HAMMOND, H. A., L. JIN, Y. ZHONG, C. T. CASKEY and R. CHAKRABORTY, 1994 Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am. J. Hum. Genet. **55:** 175–189.

HARTL, G. B., R. GÖLTENBOTH, M. GRILLITSCH and R. WILLING, 1988 On the biochemical systematics of the Bovini. Biochem. Syst. Ecol. **16:** 575–579.

KULKE, H., and D. ROTHERMUND, 1990 *A History of India.* Routledge, London.

LHOSTE, P., 1991 Cattle genetic resources of West Africa, pp. 73–89 in *Cattle Genetic Resources,* edited by C. G. HICKMAN. Elsevier, Amsterdam.

LOFTUS, R. T., D. E. MACHUGH, D. G. BRADLEY, P. M. SHARP and E. P. CUNNINGHAM, 1994a Evidence for two independent domestications of cattle. Proc. Natl. Acad. Sci. USA **91:** 2757–2761.

LOFTUS, R. T., D. E. MACHUGH, L. O. NGERE, D. S. BALAIN, A. M. BADI *et al.,* 1994b Mitochondrial genetic variation in European, African and Indian cattle populations. Anim. Genet. **25:** 265–271.

MACHUGH, D. E., R. T. LOFTUS, D. G. BRADLEY, P. M. SHARP and E. P. CUNNINGHAM, 1994 Microsatellite DNA variation within and among European cattle breeds. Proc. R. Soc. Lond. Ser. B **256:** 25–31.

MANWELL, C., and C. M. A. BAKER, 1980 Chemical classification of cattle. 2. Phylogenetic tree and specific status of the Zebu. Anim. Blood Grps Biochem. Genet. **11:** 151–162.

MASON, I. L., 1951 *The Classification of West African Livestock.* Commonwealth Bureau of Animal Breeding, Edinburgh.

MASON, I. L., and J. P. MAULE, 1960 *The Indigenous Livestock of Eastern and Southern Africa.* Commonwealth Bureau of Animal Breeding, Edinburgh.

MAULE, J. P., 1990 *The Cattle of the Tropics.* University of Edinburgh, Edinburgh.

MEADOW, R. H., 1984 Animal domestication in the Middle East: a view from the Eastern margin, pp. 309–337 in *Animals and Archaeology: 3. Early Herders and Their Flocks,* edited by J. CLUTTON-BROCK and C. GRIGSON. Oxford University Press, Oxford.

MEADOW, R. H., 1993 Animal domestication in the Middle East: a revised view from the Eastern Margin, pp. 295–320 in *Harappan Civilization,* edited by G. POSSEHL. Oxford & IBH, New Delhi, India.

NAIK, S. N., 1978 Origin and domestication of zebu cattle (*Bos indicus*). J. Hum. Evol. **7:** 23–30.

NEEL, J. V., 1973 "Private" genetic variants and the frequency of mutation among South American Indians. Proc. Natl. Acad. Sci. USA **70:** 3311–3315.

NEI, M., 1972 Genetic distance between populations. Am. Nat. **106:** 283–292.

NEI, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

NOWAK, R. M., 1991 *Walker's Mammals of the World.* The John Hopkins University Press, Baltimore.

PAMILO, P., and M. NEI, 1988 Relationships between gene trees and species trees. Mol. Biol. Evol. **5:** 568–583.

PAYNE, W. J. A., 1970 *Cattle Production in the Tropics.* Longman, London.

PAYNE, W. J. A., 1991 Domestication: a forward step in civilization, pp. 51–72 in *Cattle Genetic Resources,* edited by C. G. HICKMAN. Elsevier, Amsterdam.

RAYMOND, M., and F. ROUSSET, 1995a GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. J. Hered. **86:** 248–249.

RAYMOND, M., and F. ROUSSET, 1995b An exact test for population differentiation. Evolution **49:** 1280–1283.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SCHERF, B. D., 1995 *World Watch List for Domestic Animal Diversity.* Food and Agriculture Organization of the United Nations, Rome.

SHRIVER, M. D., L. JIN, R. CHAKRABORTY and E. BOERWINKLE, 1993 VNTR allele frequency distributions under the Stepwise Mutation Model: a computer simulation approach. Genetics **134:** 983–993.

SHRIVER, M. D., L. JIN, E. BOERWINKLE, R. DEKA, R. E. FERRELL *et al.*, 1995 A novel measure of genetic distance for highly polymorphic tandem repeat loci. Mol. Biol. Evol. **12:** 914–920.

SUZUKI, R., S. J. KEMP and A. J. TEALE, 1993 Polymerase chain reaction analysis of mitochondrial DNA polymorphism in N'Dama and Zebu cattle. Anim. Genet. **24:** 339–343.

TEALE, A. J., J. WAMBUGA, P. S. GWAKISA, G. STRANZINGER, D. BRADLEY *et al.*, 1995 A polymorphism in randomly amplified DNA that differentiates the Y chromosomes of *Bos indicus* and *Bos taurus.* Anim. Genet. **26:** 243–248.

VALDES, A. M., M. SLATKIN and N. B. FREIMER, 1993 Allele frequencies at microsatellite loci: the Stepwise Model revisited. Genetics **133:** 737–749.

VON GRAML, R., G. OHMAYER, F. PIRCHNER, L. ERHARD, J. BUCHBERGER *et al.*, 1986 Biochemical polymorphism in Egyptian Baladi cattle and their relationship with other breeds. Anim. Genet. **17:** 61–76.

WENDORF, F., and R. SCHILD, 1994 Are the early Holocene cattle in the Eastern Sahara domestic or wild? Evol. Anthropol. **3:** 118–128.

WORKMAN, P. L., B. S. BLUMBERG and A. J. COOPER, 1963 Selection, gene migration and polymorphic stability in US White and Negro populations. Am. J. Hum. Genet. **15:** 429–437.

WU, C-I., 1991 Inferences of species phylogeny in relation to segregation of ancient polymorphisms. Genetics **127:** 429–435.

Communicating editor: M. SLATKIN