

Mitochondrial diversity and the origins of African and European cattle

DANIEL G. BRADLEY, DAVID E. MACHUGH, PATRICK CUNNINGHAM, AND RONAN T. LOFTUS

Department of Genetics, Trinity College, Dublin 2, Ireland

Communicated by Fred Wendorf, Southern Methodist University, Dallas, TX, January 16, 1996 (received for review, November 3, 1995)

ABSTRACT The nature of domestic cattle origins in Africa are unclear as archaeological data are relatively sparse. The earliest domesticates were humpless, or *Bos taurus*, in morphology and may have shared a common origin with the ancestors of European cattle in the Near East. Alternatively, local strains of the wild ox, the aurochs, may have been adopted by peoples in either continent either before or after cultural influence from the Levant. This study examines mitochondrial DNA displacement loop sequence variation in 90 extant bovines drawn from Africa, Europe, and India. Phylogeny estimation and analysis of molecular variance verify that sequences cluster significantly into continental groups. The Indian *Bos indicus* samples are most markedly distinct from the others, which is indicative of a *B. taurus* nature for both European and African ancestors. When a calibration of sequence divergence is performed using comparisons with bison sequences and an estimate of 1 Myr since the *Bison/Bos Leptobos* common ancestor, estimates of 117–275,000 B.P. and 22–26,000 B.P. are obtained for the separation between Indians and others and between African and European ancestors, respectively. As cattle domestication is thought to have occurred approximately 10,000 B.P., these estimates suggest the domestication of genetically discrete aurochsen strains as the origins of each continental population. Additionally, patterns of variation that are indicative of population expansions (probably associated with the domestication process) are discernible in Africa and Europe. Notably, the genetic signatures of these expansions are clearly younger than the corresponding signature of African/European divergence.

Cattle have had an intimate and formative association with human civilization. In historic and current societies they have fulfilled key agricultural, economic, cultural, and even religious roles. The domestication of the wild ox or aurochs (*Bos primigenius*) some 10,000 years B.P. was one of the most significant achievements of neolithic peoples. Aurochsen were, from contemporary accounts (1) formidable animals and, in addition to food produce, their harnessing would have provided the first powerful source of traction to early agricultural communities.

The aurochs ranged over large tracts of the old world, including much of Asia, Europe, and North Africa. It is now extinct, having died out in most regions around 2000 years ago but reputedly surviving into medieval times in central Europe (2, 3). A prevailing view has been that all modern cattle breeds have their roots in the domestication centers of Western Asia (2–6), but this is an opinion that may be an artifact of the history of archaeological exploration itself (6).

A previous survey of mtDNA variation (7), combined with the interpretation of early neolithic faunal remains in Baluchistan (6), argue strongly for a separate origin for the cattle populations of the Indian subcontinent. The modern (as well

as the earliest) domestic cattle of this region are of the humped subspecies (whereas the binomial forms *B. primigenius*, *Bos taurus*, and *Bos indicus* are used, the data presented and inter-fertility indicate that they constitute a single species), *B. indicus*, whereas those of Europe are humpless or of *B. taurus* type.

In Africa the origins of modern cattle populations remain controversial. The first bovine domesticates of Africa are believed, from prehistoric artistic representations, to have been *B. taurus* (taurine) in morphology, and the *B. indicus* (zebu) breeds, which now predominate, entered the continent some few millennia later (4, 8). Differences between the indigenous African taurine breeds that survive and the cattle of Europe include economically important traits such as heat and disease tolerance (9). However, overall differences between European and African taurines are not as marked as those between either of these and zebu animals.

A lack of morphological divergence and the comparatively sparse finds of oxen remains in early African cultural contexts has resulted in the widely accepted assumption of a single common origin for African and European taurines in the early domestic centers of the Near East. Possible local origins for indigenous breeds in a fertile ancient North Africa are often discussed, but archaeological evidence is lacking (10–12). Particularly, no sites have been described with the types of temporal transition in faunal remains that may securely be identified with the domestication process (13). The question of an indigenous bovine African origin indeed labors under a “dearth of data and a surfeit of models” (10).

In this study we present analysis of a large data set comprising mtDNA displacement loop (D loop) sequences from 90 contemporary cattle sampled on three continents. The patterns of genetic variation revealed suggest ancient population expansions that are consistent with the demographics of the domestication process and suggest a predomestic separation for the ancestors of African and European bovines.

MATERIALS AND METHODS

Sample Collection. DNA was extracted from fresh blood collected from 13 cattle breeds selected from three continents. Indian samples, from the Tharparkar, Sahiwal, and Hariana breeds, were collected from research herds at the National Institute for Animal Genetics (Karnal, Haryana State, India). Two East African breeds, Butana and Kenana, were sampled at the National Dairy Research Center (Shukaba, Wad Medani, Sudan), and two West African breeds, N’Dama and White Fulani were accessed at the University of Ibadan (Nigeria). The six European breeds included were: Aberdeen Angus, Hereford, Jersey, Charolais, Simmental, and Friesian. These were sampled from pure bred stock kept in Irish artificial insemination centers and private herds. Seven animals were taken from each breed with the exception of Tharparkar (six). In all cases, efforts were made, using both

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: D loop, displacement loop.

Data deposition: The sequence reported in this paper has been deposited in the GenBank data base (accession no. U51806–U51842).

pedigrees and the knowledge of local herdsmen, to ensure that animals were not closely related. D loop sequences from two samples from each breed mentioned have been included in a previous publication (7). Bison sequences were obtained from GenBank (R. N. Beech, J. Sheraton, R. Polziehn, and C. Strobeck, personal communication) (accession numbers: BBU12936, BBU12946, BBU12948, BBU12955, and BBU12959).

Amplification and Sequencing. mtDNA was isolated from blood using the method of Lindberg *et al.* (14). D loops were amplified using the polymerase chain reaction with primers constructed using the published proline tRNA (5'-CTGCA-GTCTACCATCAACC-3') and 12S rRNA (5'-CTCCTCG-GACAAGATATTAG-3') gene sequences (15). Amplifications and product purifications were carried out as described previously (7). Standard double-stranded DNA sequencing was performed using approximately 250–500 ng of amplification product and the following primers: 5'-GTACATAACA-TTAATGTAAT-3'; 5'-AAACCAGCAACCCGCT-3'.

Sequence Analysis. Alignment of sequences was achieved using the CLUSTALW package (16). Sites representing a gap in any of the aligned sequences were excluded from the analysis, and distances between whole D loop sequences were estimated using the substitution model of Tamura and Nei (17) with a γ distribution parameter value $\alpha = 0.11$ (an estimate from human whole D loop data). This procedure is incorporated in the MEGA package (18). A minimum spanning tree was constructed by hand after the recommendations of Excoffier and Smouse (19), and additional phylogenies (not shown) were constructed using the neighbor-joining (20) and maximum likelihood algorithms incorporated in the PHYLIP package (21). Analysis of molecular variance (AMOVA) was performed using software provided by Excoffier *et al.* (22). Pairwise genetic difference analyses were performed using programs supplied by Rogers (23–25).

RESULTS

Variation in the Bovine mtDNA D Loop. Twenty-six almost complete bovine D loop sequences have been reported and analyzed previously (7) (GenBank accession nos. L27712–L27737). Here, these were aligned with recently published bison sequences (bases numbered 15816 to 354) to give a novel calibration of the molecular clock in the whole D loop. In addition, primers were designed to analyze 370 bp from the most variable region (bases numbered 16032 to 63) (7), and this portion was sequenced in another 64 individuals. Analysis of the total 90 partial bovine D loop sequences are discussed below.

Alignment of these 90 370-bp sequences with one bison illustrated 55 unique *Bos* haplotypes (Figure 1). In 42 assayed European cattle, 20 haplotypes were found, including one (*ANI*) that was represented 16 times. Two other haplotypes were shared between three animals each, and three were found in duplicate. In African breeds, 19 unique sequences were encountered in 28 mtDNA chromosomes, including two doubles and one (*ND4*) that was repeated eight times. 20 sequences of Indian origin included 16 haplotypes, two of which were repeated twice and one that was represented three times.

No mtDNA haplotypes were shared by animals from different continents, but in many cases, genomes assayed from different breeds within continents yielded identical sequences. Notably, the most common African sequence was encountered in animals of both *B. indicus* (five) and *B. taurus* (three) morphologies.

Differences Between Haplotypes. The nucleotide variations between the 55 unique haplotypes are illustrated in Fig. 1. Eighty variable sites included one 7-bp deletion found only in bison and three 1-bp insertion/deletions. Only one substitution observed within cattle sequences and four between cattle and bison were the result of nucleotide transversions, reflecting the heavy transitional bias that has previously been described for both bovine and human control regions (7, 26). The geographic

structuring of sequence variation is observable from the large number of consistent nucleotide differences observed between the Indian samples and other bovines and also, to a lesser extent, from the polymorphisms that are typical of animals of African rather than European origin.

Global Genetic Structure. Genetic structuring was investigated using the AMOVA method (22). This procedure uses information from both the estimated divergence between haplotypes and the frequencies at which each is represented in a population grouping. Through estimating variance components, the structuring of genetic variation between different hierarchical levels may be assessed. A global AMOVA, using 90 bovine 370-bp sequences, estimated that 84% of the variance could be accounted for by the three continental divisions.

The integrity of the African/European division was examined more closely by excluding the Indian breeds from the analysis. In this case, the single continental division accounted for 51% of the variance. The within-continent, between-breed figure (<4%) was eclipsed by the variation at the individual level (45%), despite the presence of very distinct zebu and taurine breeds in the African sample. mtDNA variation seems to be a poor assay for genetic relationships at the breed level. In order to test the genetic validity of the geographical division into African and European populations, 1000 random divisions of the data matrix were generated and compared with the actual partition (19). It was confirmed as belonging to a class of <1% of permutations, which explained a maximum amount of the molecular variance.

Phylogenetic Tree Construction. In order to focus on the relationship between African and European populations, a minimum spanning tree was constructed using only the 370-bp sequences from these two continents (Figure 2). This network allows extant sequences as internal nodes, and the frequency with which each haplotype occurs is indicated by the area of the circle representing it. The graph topology is strikingly bipolar, with each of the two continental groups clustering separately around one of two predominant and possibly ancestral sequences. These two putative ancestral haplotypes are separated by three substitutions, and the starlike patterns surrounding each may be argued as being indicative of two separate population expansions; a possibility discussed in more detail below. The tree shown is one of several, equally valid alternatives, each of which display these same major topological features and which were also observed in neighbor-joining and maximum-likelihood summaries of the same data set. As in a previous study (7), mtDNA haplotypes showed no tendency to cluster into breed groups, and no distinction was discernible between African taurine and zebu sequences. When considered jointly with West African studies of microsatellite and Y chromosome variation (27, 28), these data reveal African *B. indicus* as hybrids with ancient taurine maternal lineages in a predominantly zebu genetic background.

Divergence Times Between Continental Populations. The *Bison/Bos* split is estimated to have occurred at least one million years ago (7, 29). This allowed the calibration of the D loop clock in two ways. Firstly, a calibration for the 370-bp hypervariable region was performed after Vigilant *et al.* (30). Two out of a total of 116 substitutions that have been inferred overall in the bovine D loop are transversions, giving an estimate of the transition/transversion ratio of 57:1. Four transversions are detectable between bison and all cattle sequences, and correspondingly it is estimated that 228 transitions have occurred in the 363-bp segment compared in the 1 Myr since the most recent common ancestor. This leads to a two-lineage rate estimate of 62.8% divergence per Myr or, alternatively, the accumulation of 1-bp substitution in a 370-bp fragment per 4,303 years.

Second, comparisons between previously published, whole D loop sequences were utilized. Estimations of the number of nucleotide substitutions within the D loop are complicated by

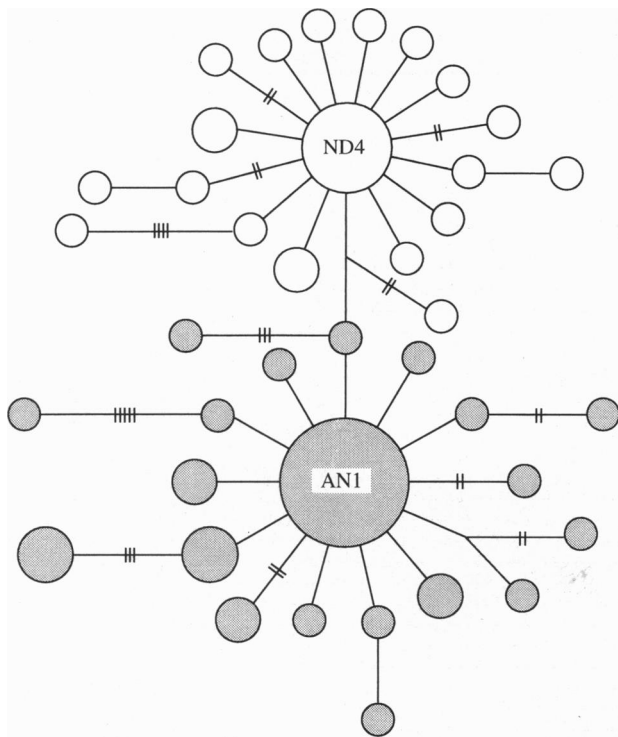


FIG. 2. Minimum spanning tree constructed from 42 European (shaded circles) and 28 African (open circles) 370-bp partial D loop sequences. The number of times each variant is represented is proportional to the area of its circle, and the bulk of variants are represented only once. The number of substitutions between sequences is singular, except when two or more differences are denoted by the corresponding number of cross-hatched lines. Only the two topologically and numerically predominant variants are named: ND4 (African) and AN1 (European). Several of the African nodes present equally parsimonious alternatives for a tree root (not shown).

Europe comparisons in Fig. 3 display curves indicative of expansions in the recent evolutionary past. Using estimators developed by Rogers (25) and the 370-bp region evolutionary rate detailed above, the time of the African expansion is given as 9000 B.P., and that in Europe is given as 5000 B.P. Notably, the distribution from African versus European sequence pairwise comparisons unambiguously leads each of the within-continent distributions, indicating strongly that the separation of the ancestral continental populations occurred prior to major expansions that might be closely linked to domestication events.

The whole D loop data set (7) is more limited in sample size (only eight and 12 sequences per continent) but yields results

Table 1. Estimated time depths for the common ancestors of continental strains

	Time (years B.P.) (Divergence)	
	Total D loop	Partial D loop
Bison vs. Bos	1,000,000 (0.3009)	1,000,000 (0.6281)
India vs. Others	275,000 (0.0827)	117,000 (0.0738)
Africa vs. Europe	26,000 (0.0079)	22,000 (0.0138)

Corresponding mean sequence divergence estimates are given in brackets. Estimates are derived from a figure of 1 Myr for the *Bison/Bos* separation and employ (a) near-complete D loop sequences and the Tamura and Nei (18) substitution model; (b) the larger partial D loop dataset with a calibration based on the number of transversions observed between bison and cattle with no correction employed for comparisons between cattle.

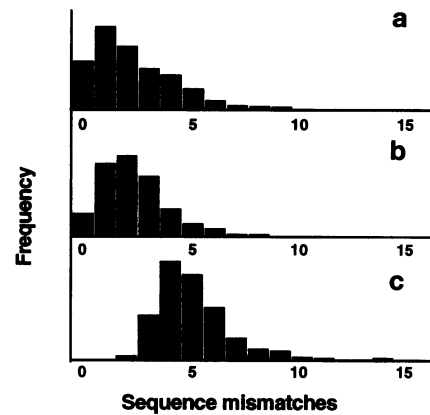


FIG. 3. Frequency distributions of the number of sequence differences observed in pairwise comparisons. The curves correspond to 370-bp D loop sequence comparisons, and the number of mismatches are given on the horizontal axis with relative frequency of each category represented on the vertical scale. These derive from comparisons within the European sample (a), comparisons within the African sample (b), and intercontinental comparisons between African and European sequences (c). In each case smooth, single-peaked curves are apparent. The distribution resulting from the between-continent comparison clearly leads those of within-continent derivation.

of similar magnitude under the same analysis. Corresponding expansion time estimates for European and African populations are 9000 and 11,000 B.P., respectively.

DISCUSSION

Commentators on the prehistory of cattle herding on the African continent are consistent in the assertion that the earliest domesticates were taurine and that the zebu animals, which now predominate are the results of later migrations (4, 7, 8). However, there is uncertainty, primarily from lack of evidence, as to whether the initial herds were the product of local aurochs domestication or whether they share a common origin with European cattle in the early domestic centers of the Near East.

The integrity of a biological partition that corresponds to the geographical division between African and European sequences is attested to by several analyses. The actual division gives a significantly nonrandom partition of the molecular variance, and three different methods of phylogeny construction separate sequences into their continental clusters (Fig. 2). Additionally, no duplications of haplotypes are observed between the two continents, whereas many are observed within.

In each continent, the most frequent sequence also forms the center of a radiation of variants which differ from it by one or only a few substitutions (Fig. 2). This numerical and topological predominance suggests that these two, *AN1* and *ND4*, may represent separate ancestral mitochondrial types for European and African cattle, respectively.

Distributions of pairwise differences provide an alternative visualization of patterns of variation, and here the close relationships of sequences within either Africa or Europe are illustrated by narrow curves with modal values of one or two mismatches between sequences (Fig. 3). In simulation studies, single peaked, smooth mismatch distributions have been noted to be typical of populations that have undergone past expansions from relatively narrow population bases (24).

The estimates of past expansion times of African and European cattle populations are 9000 and 5000 B.P., respectively. These estimates are subject to wide and indeterminate errors but may be viewed as roughly consistent with demographic events corresponding to the domestication process. The neolithic transition led to major expansions of human populations,

which are reflected in modern genetic geography (33), and it is to be expected that the development and subsequent success of cattle herding would have resulted in a profound expansion in the domesticate populations from an initially narrow base.

The mismatch frequency graph that results from comparisons between African and European mitochondrial variants markedly leads the two within-continent curves by approximately three mutational units (Fig. 3). This suggests that the ancestral separation of the continental progenitors markedly preceded domestication: an assertion that is independent of substitution rate estimates. Moreover, using calibrations based on both total and partial D loop sequence comparisons with bison and a conservative 1 Myr dating of its divergence from the bovine ancestor, the mean sequence divergence between Africa and Europe yields estimates of at least 22,000 years since the existence of a common ancestor for the two continental strains. A wide error is associated with substitution rate estimates in the D loop, but some security is derived by the similarity in magnitude of our figures and the range of estimates derived using a range of methods from equivalent human sequence data (17, 30–32).

The presence of ancient cultural debris and rock engravings and paintings in the harshest of modern Saharan environments are indicators of past eras when herders and their flocks inhabited a more fertile North Africa (8, 34). A wet climatic phase between 10,000 and 8000 B.P. affected all of north and east Africa and may have incorporated the first cattle domestication. Sites in Nabta Playa and Bir Kiseiba in eastern Sahara have yielded putative *Bos* bones dated up to 9000 B.P. (35). It is argued that the dry contemporary climate at these sites was simultaneously one in which cattle would have existed only with human intervention and also one that may have predisposed the early Neolithic communities to pastoralism (36). The tentative placing of these Eastern Saharan remains in a domestic context would be a strong indication of a local, early domestication with a possibility of independence from Near Eastern centers. However this conclusion is regarded by some as insecure (11, 12).

The wild progenitors of sheep and goats are not native to Africa, and their appearance in the Saharan faunal record post 7700 B.P. (36) is an indication of the presence of a herding culture that had some roots in the Levant. The oldest cattle remains in an obvious domestic context have been found along with ovicaprid remains in Capéletti, Algeria, and date to 6530 B.P. (37).

The levels and patterns of mitochondrial sequence diversity uncovered in this study do not point toward a simple model of a single Near Eastern origin for African and European cattle within the 10,000 year time frame of domestic history. The possibility may be argued that two divergent lineages coexisted in a single ancestral domestic population and that differential loss of these occurred in two daughter groups, but this represents the most labored interpretation of the genetic data. Alternatively, the biological separation observed could be the result of adoption of local wild oxen into existing European or African herds by early herders. However, the evidence is most suggestive of two domestic origins that were either temporally or spatially separate and that involved divergent strains of taurine progenitors. This is consistent with a Near Eastern origin for European cattle and an African origin for the breeds of that continent.

The dating of the putative African bovine population expansion, although comprising a rough estimate, seems older than that deduced in European patterns of variation. This provides some tentative support for an earlier and possibly Saharan domestication process that may have been independent of the later Near Eastern influences, which are detectable through the presence of ovicaprid herding.

We acknowledge the assistance of the following in the collection of blood samples: D. S. Balain and V. J. Shankar (NIAG, Karnal,

Haryana, India); L. O. Ngere (University of Ibadan, Nigeria); A. M. Badi, J. Rizgalla, and A. Hassan (APRA, Khartoum, Sudan). We also thank P. M. Sharp, A. T. Lloyd, C. M. Meghen, C. S. Troy, and the Irish National Centre for Bioinformatics for valuable assistance. This work was funded by the European Commission DGXII Science and Technology in Developing Countries Program.

1. Isaac, E. (1962) *Science* **137**, 195–204.
2. Payne, W. J. A. (1991) *Cattle Genetic Resources*, ed. Hickman, C. G., (Elsevier Science), World Animal Series B7, 51–72.
3. Epstein, H. & Mason, I. L. (1984) in *Evolution of Domesticated Animals*, ed. Mason, I. L. (Longman Inc., New York), pp 6–27.
4. Epstein, H. (1971) *The Origin of the Domestic Animals of Africa* (Africana Publishing), Vol 1, 185–555.
5. Perkins, D., Jr. (1969) *Science* **164**, 177–179.
6. Meadow, R. H. (1993) in *Harappan Civilization*, 2nd edition, ed. G. Possehl. (New Delhi, Oxford and IBH) pp. 295–320.
7. Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M. & Cunningham, E. P. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2757–2761.
8. Muzzolini, A. (1983) *L'Art Rupestre du Sahara Central: Classification et Chronologie. Le Bœuf dans la Préhistoire Africaine*. Ph.D. thesis. (Univ. of Provence, Aix-Marseille).
9. Murray, M., Morrison, W. I. & Whitelaw, D. D. (1982) *Adv. Parasitol.* **21**, 1–68.
10. Gautier, A. (1987) in *Prehistory of Arid North Africa, Essays in Honor of Fred Wendorf*. ed. A. E. Close (Southern Methodist University Press, Dallas, TX), pp. 163–180.
11. Clutton-Brock, J. (1989) in *The Walking Larder: Patterns of Domestication, Pastoralism, and Predation*. ed. Clutton-Brock, J. (Allen and Unwin, London), pp. 200–206.
12. Smith, A. B. (1986) *Afr. Archaeol. Rev.* **4**, 197–203.
13. Meadow, R. H. (1989) in *The Walking Larder: Patterns of Domestication, Pastoralism, and Predation*. ed. Clutton-Brock, J. (Allen and Unwin, London), pp. 80–10.
14. Lindberg, G. L., Koehler, C. M., Mayfield, J. E., Myers, A. M. & Beitz, D. C. (1992) *Biochem. Genet.* **30**, 27–33.
15. Anderson, S., DeBruijn, M. H. L., Coulson, A. R., Eperon, I. C., Sanger, F. & Young, I. G. (1982) *J. Mol. Biol.* **156**, 683–717.
16. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
17. Tamura, K. & Nei, M. (1993) *Mol. Biol. Evol.* **10**, 512–526.
18. Kumar, S., Tamura, K. & Nei, M. (1993) *MEGA: Molecular evolutionary genetics analysis, version 1.0*. Pennsylvania State University, State Park, PA.
19. Excoffier, L. & Smouse, P. E. (1994) *Genetics* **136**, 343–359.
20. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
21. Felsenstein, J. (1993) *Phylip (Phylogeny Inference Package) Version 3.5c*. University of Washington, Seattle, WA.
22. Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) *Genetics* **131**, 479–491.
23. Rogers, A. R. & Harpending, H. (1992) *Mol. Biol. Evol.* **9**, 552–569.
24. Harpending, H. C., Sherry, S. T., Rogers, A. R. & Stoneking, M. (1993) *Curr. Anthropol.* **34**, 483–496.
25. Rogers, A. R. (1995) *Evolution* **49**, 608–615.
26. Horai, S., Kondo, R., Murayama, K., Hayashi, S., Koike, H. & Nakai, N. (1991) *Philos. Trans. R. Soc. Lond.* **333**, 409–417.
27. Meghen, C., MacHugh, D. E. & Bradley, D. G. (1994) *World Anim. Rev.* **78**, 59–66.
28. Bradley, D. G., MacHugh, D. E., Loftus, R. T., Sow, R. S., Hoste, C. H. & Cunningham, E. P. (1994) *Anim. Genet.* **25**, 7–12.
29. Hartl, G. B., Goltenboth, R., Grillitsch, M. & Willing, R. (1988) *Biochem. Syst. Ecol.* **16**, 575–579.
30. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. (1991) *Science* **253**, 1503–1507.
31. Kolman, C. J., Bermingham, E., Cooke, R., Ward, R. H., Arias, T. D. & Guionneau-Sinclair, F. (1995) *Genetics* **140**, 275–283.
32. Lundstrom, R., Tavaré, S. & Ward, R. H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 5961–5965.
33. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ), pp. 290–301.
34. Smith, A. B. (1992) *Annu. Rev. Anthropol.* **21**, 125–141.
35. Gautier, A. (1984) in *Prehistory of the Eastern Sahara*. ed. Close, A. E. (Academic Press, New York), pp. 317–344.
36. Wendorf, F. & Schild, R. (1994) *Evol. Anthropol.* **3**, 118–128.
37. Roubet, C. (1978) *Anthropologie* **82**, 583–586.