

Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo

S. Kumar, M. Nagarajan, J. S. Sandhu, N. Kumar, V. Behl and G. Nishanth

Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500007, India

Summary

Water buffalo (*Bubalus bubalis*) is broadly classified into river and swamp categories, but it remains disputed whether these two types were independently domesticated, or if they are the result of a single domestication event. In this study, we sequenced the mitochondrial D-loop region and *cytochrome b* gene of 217 and 80 buffalo respectively from eight breeds/locations in northern, north-western, central and southern India and compared our results with published Mediterranean and swamp buffalo sequences. Using these data, river and swamp buffalo were distinguished into two distinct clades. Based upon the existing knowledge of cytogenetic, ecological and phenotypic parameters, molecular data and present-day distribution of the river and swamp buffalo, we suggest that these two types were domesticated independently, and that classification of the river and swamp buffalo as two related subspecies is more appropriate.

Keywords divergence, domestication, mitochondrial DNA, river buffalo, swamp buffalo.

Introduction

The domestic water buffalo is a species of paramount importance to many Asian countries, particularly to those in the Indian subcontinent. This species is a major source of milk, meat, draft power, hide and employment to marginal farmers and landless labourers in many Asian countries. In spite of the tremendous economic importance of buffalo, interest in buffalo genomics has been limited (Moore *et al.* 1995; Barker *et al.* 1997; Navani *et al.* 2002; Iannuzzi *et al.* 2003; Kierstein *et al.* 2004; Kumar *et al.* 2006).

The domestic water buffalo are broadly classified into two major categories based upon their phenotype, behaviour and karyotype: river buffalo ($2n = 50$) found in the Indian subcontinent, Middle East and Eastern Europe and swamp buffalo ($2n = 48$) distributed in China, Bangladesh, the Southeast Asian countries and north-eastern states of India (Cockrill 1981). The time and place of domestication of this

species has not yet been resolved. It is believed that buffalo was domesticated in the Indus civilization, the Yangtze and the Euphrates and Tigris some 5000 years ago (Cockrill 1981). However, Chen & Li (1989) proposed that the domestication of this species occurred in China during the fifth millennium BC. Two contradictory suggestions have been proposed: that these two types were domesticated independently (Tanaka *et al.* 1996; Lau *et al.* 1998) and that these are a product of a single domestication event (Kierstein *et al.* 2004). Kierstein *et al.* (2004) analysed the mitochondrial (mt) D-loop DNA sequences of Indian river and swamp (Carabao) buffalo from Brazil and Mediterranean animals from Italy. They suggested that the river and swamp buffalo were the product of a single domestication event probably in the Indian subcontinent. Earlier, Kikkawa *et al.* (1997) suggested that the domestic river and swamp lines arose independently based on differences in their *cytochrome b* sequences. Although India is a hotspot of buffalo biodiversity (George *et al.* 1988), and it is generally believed that buffalo were domesticated in India (Cockrill 1981), none of the studies dealing with buffalo domestication (Kikkawa *et al.* 1997; Lau *et al.* 1998; Kierstein *et al.* 2004) sampled animals from Indian locations. Here we report the sequence analysis of the mt D-loop region ($n = 217$) and the *cytochrome b* gene ($n = 80$) in Indian river buffalo and compare them with published sequences of 10

Address for correspondence

Dr Satish Kumar, National Facility for Transgenic and Gene Knockout Mice, Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India.
E-mail: satishk@ccmb.res.in

Accepted for publication 29 January 2007

D-loop and seven *cytochrome b* gene haplotypes from swamp buffalo. Our analyses provide strong evidence for the independent origin of domestic river and swamp buffalo.

Materials and methods

Sample collection

Field visits were arranged to the breeding tracts of river buffalo breeds in northern, north-western, central and southern India. Blood samples were collected from 217 animals representing eight breeds from eight locations. Two to three unrelated animals with typical breed characteristics were sampled from a given hamlet. Fresh blood was obtained by puncturing the jugular vein and DNA was isolated using the extraction method of Sambrook & Russell (2001).

Amplification and sequencing of mtDNA

To amplify the buffalo mtDNA D-loop region, we designed two overlapping sets of primers, HVR-I (5'-GGAAGAAAC-TGCAGTCTCACCAT-3', 5'-ATGGGCTGATTAGCCATTAGTCC-3') and HVR-II (5'-GGTTCTTCTTCAGGGCCATCTC-3', 5'-CAGTGCCTTGCTTTGGGTTAAGC-3'), from the available *Bos taurus* mtDNA sequence (Anderson *et al.* 1982). The *cytochrome b* gene was amplified by PCR using two overlapping sets of primers, Cyto-I (5'-ATCATTCTGAGGG-GCAACAG-3', 5'-CTTCAGCTTTGGGTGTTGGT-3') and Cyto-II (5'-GACAATATCCGAAAATCCCAC-3', 5'-TTCACCTCAACCAGACTAGTACC-3'), from the *B. taurus* mtDNA sequence (Anderson *et al.* 1982).

After PCR using standard conditions, sequencing was performed using Big Dye and carried out on ABI 3700 and 3730 automated DNA sequencers (Applied Biosystems).

Phylogenetic and reduced median network analyses

The DNA sequences were edited manually using AUTOASSEMBLER (Perkin Elmer) and the CLUSTALX program (Thompson *et al.* 1997) was used for multiple alignments. We compared 135 D-loop haplotypes of a 945-bp fragment obtained from 217 river buffalo from India (EF464323–EF464457) to 11 haplotypes of Mediterranean (river) buffalo and 10 haplotypes of swamp (Carabao) buffalo reported by Kierstein *et al.* (2004). *Bos taurus* sequence (NC_006853) was used as the out-group. A Bayesian phylogenetic tree was constructed by Markov chain Monte Carlo (MCMC) method as implemented in the MRBAYES 3.1 package (Ronquist & Huelsenbeck 2003) using the general time reversible substitution model with the invariant site plus eight gamma categories. Initial runs were performed to find out the fluctuating value of the likelihoods of Bayesian trees and stationarity was reached at 500 000 generations. Analyses of MCMC chains were run for 3×10^6 cycles and

the trees were sampled every 100 generations. All the sampled trees preceding stationarity were discarded (burnin = 5000) and the remaining tree samples were used to generate a 50% majority rule consensus tree. Analyses were repeated four times. Similarly, a maximum parsimony (MP) tree was computed using MEGA 3.1 (Kumar *et al.* 2004). The close neighbour-interchange algorithm was used with a search level of three. The searches included 100 replications of random addition trees and the robustness of nodes was confirmed by 1000 bootstrapping. A Bayesian phylogenetic tree was also constructed from four Indian river buffalo *cytochrome b* haplotypes (EF409939–EF409942) and seven swamp buffalo haplotypes (Kikkawa *et al.* 1997) essentially as described above except that the MCMC chains were run for 1×10^6 cycles (burnin = 2500). The *B. taurus* sequence (NC_006853) was used as the out-group. Reduced median network was applied to the dataset containing 238 mtDNA D-loop sequences (921 bp) and 11 *cytochrome b* (1125 bp) haplotypes of river and swamp buffalo using NETWORK 4.1.1.2 (Bandelt *et al.* 1995). Gaps were excluded in reduced median network and parameters were set to a weight of two and threshold value of one.

Divergence time

Divergence time and time of the most recent common ancestor (TMRCA) were calculated using 375 bp of the hypervariable region of the mtDNA D-loop (nucleotides 15982–16354 of the *Bubalus bubalis* mt sequence, (AF547270). We used the MDIV program (Nielsen & Wakeley 2001), which is based on the coalescent theory and employs Bayesian inference to estimate the likelihood parameters of θ , M , T and TMRCA by calculating posterior probabilities assuming the HKY mutation model. Initial runs were conducted for different values of M (0–5) and T (0–7). Both parameters were adjusted to $M = 4$ and $T = 3$ and the process was repeated five times with different seed values. For each replicate, the settings of Markov chain were fixed at 5×10^6 steps and a burn-in time of 5×10^5 steps. The pairwise estimates of T and TMRCA were converted to time using $t = T\theta/2\mu$ and $tMRCA = TMRCA\theta/2\mu$ (Steeves *et al.* 2005), where μ is the mutation rate. The mutation rate was assumed to be 32%/Myr, based on a calibration for the bison control region (Shapiro *et al.* 2004) and a generation time of 6 years. Similarly, we obtained estimates of divergence time and TMRCA of the river and swamp buffalo from *cytochrome b* haplotypes; the mutation rate was assumed to be 2.5%/Myr (Meyer *et al.* 1990).

Results

A Bayesian phylogenetic tree constructed from the mtDNA D-loop sequences, along with posterior probabilities for individual clades, is presented in Fig. 1. This tree, rooted by

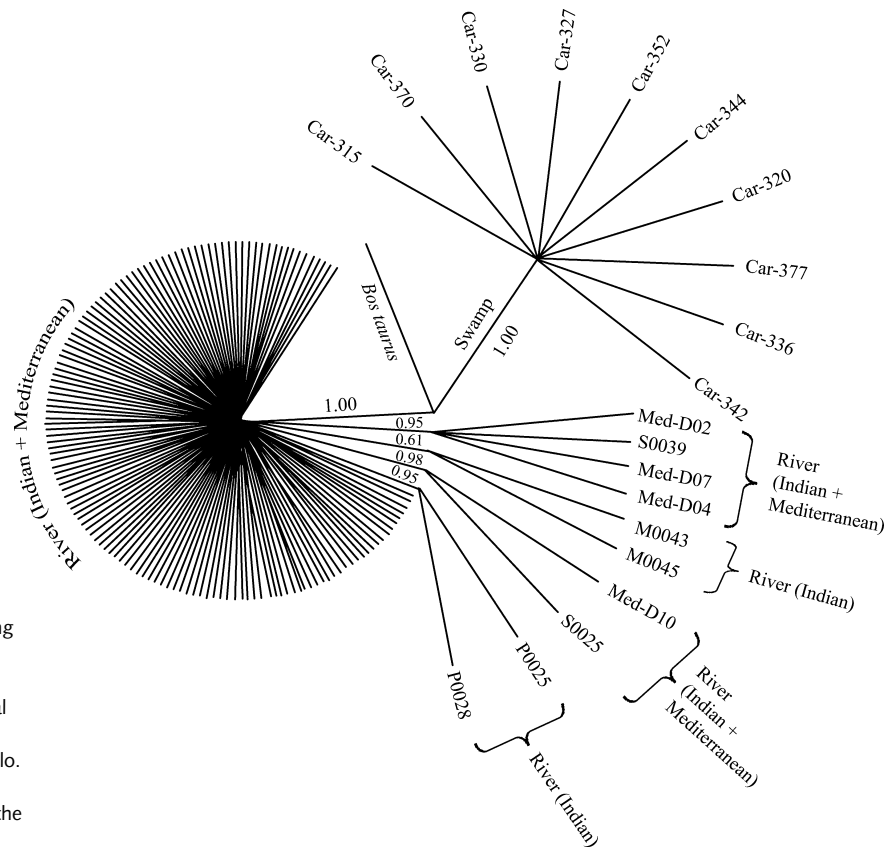


Figure 1 Bayesian phylogenetic tree showing river and swamp buffalo as distinct clades. The tree was constructed using MRBAYES 3.1 with 945-bp sequences of the mitochondrial DNA D-loop region of the Indian river, Mediterranean and swamp (Carabao) buffalo. The tree was rooted by *Bos taurus*. The numbers above the branches correspond to the respective posterior probabilities.

B. taurus, was well resolved and showed two distinct clades with a high posterior probability of 1.00. One of these two clades encompassed all the haplotypes of the Indian and Mediterranean buffalo with five internal branches. The other clade was comprised of 10 swamp haplotypes. The MP tree (Fig. S1) revealed a topology similar to the Bayesian tree except that the river clade in the parsimony tree differentiated into two additional internal branches represented by the Indian buffalo. Reduced median network (Fig. S2) showed the swamp haplotypes as a distinct group at least 62 mutation steps away from any given river haplotype. Thus, phylogenetic and network analyses separated the river and swamp buffalo into two distinct clades.

We also sequenced 458 bp of the mt *cytochrome b* gene of 80 animals (AF461785–AF461805; AF462612–AF462631; AY072746–AY072765; AY072941–AY072960) from eight recognized river breeds from eight geographical regions in India. Four *cytochrome b* haplotypes were found (Table S1). We then extended the size of the four Indian river haplotypes to 1125 bp and compared these with seven haplotypes of swamp buffalo obtained from sequence information on 32 swamp animals sampled from Japan, Taiwan, Thailand, Indonesia, Philippines and Bangladesh (Kikkawa *et al.* 1997). The Bayesian phylogenetic tree revealed two distinct clades with an extremely high posterior probability (Fig. 2). One of these two clades incorporated all four river buffalo haplotypes and the second

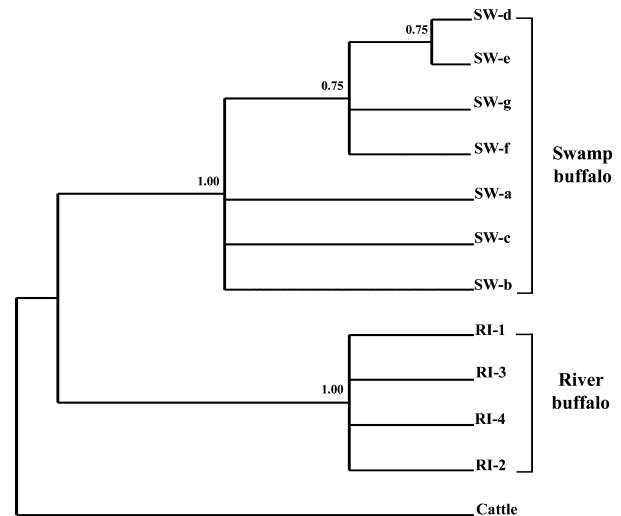


Figure 2 Bayesian phylogenetic tree showing the Indian river and swamp buffalo as two distinct clades. The tree was constructed using MRBAYES 3.1 from 1125-bp sequences of the mitochondrial DNA *cytochrome b* gene and was rooted by *Bos taurus*. Numbers above the branches correspond to the respective posterior probabilities. The swamp sequences have been described by Kikkawa *et al.* (1997).

clade had seven swamp buffalo haplotypes. The reduced median network also showed that the river haplotypes were a distinct group from the swamp haplotypes (Fig. S3). More

importantly, the three polymorphic sites observed in the Indian river buffalo were specific to these animals (Table S1). Similarly, polymorphic sites found in the swamp haplotypes were specific to the swamp animals.

We estimated the time of divergence and TMRCA between the river and swamp buffalo as 128 040 and 169 883 years respectively using 375 bp of the hypervariable region of the mtDNA D-loop. Further, we calculated a divergence time of 270 887 years from 1125 bp *cytochrome b* gene sequences assuming a mutation rate of 2.5%/Myr for this locus (Meyer *et al.* 1990).

Discussion

There have been contradictory suggestions in the literature on the origin of the domestic river and swamp buffalo (Tanaka *et al.* 1996; Lau *et al.* 1998; Kierstein *et al.* 2004). Tanaka *et al.* (1996) sequenced the *cytochrome b* gene of one animal each of the river and swamp type from Bangladesh and one wild buffalo from Sri Lanka. They estimated a divergence time of 1.7 Myr between the river and swamp buffalo and inferred that the domestic buffalo originated from at least two populations. Lau *et al.* (1998) proposed independent domestication of the river and swamp buffalo based upon the D-loop and *cytochrome b* gene sequencing, although the D-loop neighbour-joining tree showed intermingling of the river and swamp haplotypes and had low bootstrap support. The lack of resolution in the tree was probably due to the use of only 158 bp of the D-loop region and missing sequence information. Kierstein *et al.* (2004) have analysed the entire D-loop region of the river and swamp buffalo sampled from Brazil and Italy and concluded that the two types of buffalo descended from a single domestication event some 5000 years ago. There may be several reasons that the domestication events in buffalo are not conclusive. First, although India is the place of origin of all the major river buffalo breeds, previous studies (Tanaka *et al.* 1996; Kikkawa *et al.* 1997; Lau *et al.* 1998; Kierstein *et al.* 2004) did not include animals from India. Secondly, river and swamp buffalo have co-existed in locations where samples have been collected and the two types of buffalo interbreed. Thirdly, no genetic data have been reported from Chinese buffalo. In the present study, we did not sample animals from the eastern parts of India because this region along with Bangladesh is likely to have both the river and swamp animals (Kikkawa *et al.* 1997). Our *cytochrome b* gene analysis confirmed the dichotomy of the Indian river and swamp buffaloes from several Asian countries including Japan, Taiwan, Thailand, Indonesia, the Philippines and Bangladesh. These results are consistent with those reported by Kikkawa *et al.* (1997), although they used only 13 river buffalo samples from Pakistan, Bangladesh, Sri-Lanka and Italy. Swamp buffalo D-loop sequences used in our analysis had very low diversity, reflecting an

unrepresentative sampling (Kierstein *et al.* 2004), while the swamp sequences in the *cytochrome b* analysis had comparatively high diversity of 0.80% compared with 0.18% for Indian river buffalo.

The estimates of divergence time between the river and swamp buffalo in the previous studies varied from 10 000–15 000 years (microsatellite loci; Barker *et al.* 1997a) to 1.7 Myrs (the *cytochrome b* gene; Tanaka *et al.* 1996). Lau *et al.* (1998) obtained a divergence time estimate of 28 000 or 87 000 years from the D-loop sequences assuming a mutation rate of 36.6% or 11.8%/Myr respectively. The statistical approaches (Nei 1976, 1987) used by Barker *et al.* (1997) and Lau *et al.* (1998) assumed that the populations were at equilibrium for the loss of diversity through drift and gain in diversity through migration. However, the datasets in these studies were not tested for this condition. In the present study, we obtained a divergence time estimate of 128 040 years from D-loop sequences using a Bayesian approach and assuming a mutation rate of 32%/Myr (Shapiro *et al.* 2004). Similarly, we calculated a divergence time of 270 887 years from 1125 bp of the *cytochrome b* gene assuming a mutation rate of 2.5%/Myr (Meyer *et al.* 1990). The latter estimate was much lower than those obtained from the RFLP analysis of the *cytochrome b* gene reported by Amano *et al.* (1994) and Tanaka *et al.* (1995) and the *cytochrome b* gene sequencing by Tanaka *et al.* (1996) and Kikkawa *et al.* (1997).

Cytogenetic (Harisah *et al.* 1989), ecological and phenotypic parameters (Cockrill 1981); *cytochrome b* sequences (Amano *et al.* 1994; Tanaka *et al.* 1995; 1996; Kikkawa *et al.* 1997); and our results on the mtDNA D-loop region and the *cytochrome b* sequence analyses strongly support the classification of the river and swamp buffalo as two related subspecies, i.e. *B. bubalis bubalis* and *B. bubalis carabane-sis* (http://animaldiversity.ummz.umich.edu/site/accounts/information/Bubalus_bubalis.html). Loftus *et al.* (1994) reported an average sequence divergence of 7.9% between the mtDNA D-loop of *B. taurus* and *Bos indicus*, while the corresponding value for the river and swamp buffalo in our study was 8.6%. The *cytochrome b* sequence divergence between the river and the swamp buffalo was 2.6%. These differences and present-day distribution of the river and swamp buffalo suggest an independent domestication of these two types.

Acknowledgements

This work was supported by the Department of Biotechnology, Government of India, New Delhi through a part of a research grant to S.K.

References

- Amano T., Miyakoshi Y., Takada T., Kikkawa Y. & Suzuki H. (1994) Genetic variants of ribosomal DNA and mitochondrial

- DNA between swamp and river buffaloes. *Animal Genetics* **25**, 29–36.
- Anderson S., De Bruijn M.H., Coulson A.R., Eperon I.C., Sanger F. & Young I.G. (1982) Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. *Journal of Molecular Biology* **156**, 683–717.
- Bandelt H.J., Forster P., Sykes B.C. & Richards M.B. (1995) Mitochondrial portraits of human populations. *Genetics* **141**, 743–53.
- Barker J.S.F., Tan S.G., Selvaraj O.S. & Mukherjee T.K. (1997) Genetic variation within and relationships among population of Asian water buffalo (*Bubalus bubalis*). *Animal Genetics* **28**, 1–13.
- Chen Y.S. & Li X.H. (1989) New evidence of the origin and domestication of the Chinese swamp buffalo (*Bubalus bubalis*). *Buffalo Journal* **1**, 51–5.
- Cockrill W.R. (1981) The water buffalo: a review. *British Veterinary Journal* **137**, 8–16.
- George M., Balaine D.S., Vij P.K. & Kumar S. (1988) Conservation and management of buffalo genetic resources of India. In: *Buffalo Production and Health* (Ed. by R. Nagarcenkar), pp. 31–37. ICAR, New Delhi.
- Harisah M., Azmi T.I., Hilmi M., Vidyadaran M.K., Bongso T.A., Nava Z.M., Momongan V. & Basrur P.K. (1989) Identification of crossbred buffalo genotypes and their chromosome segregation patterns. *Genome* **32**, 999–1002.
- Iannuzzi L., Di Meo G.P., Perucatti A., Schibler L., Incarnato D., Gallagher D., Eggen A., Ferretti L., Cribru E.P. & Womack J. (2003) The river buffalo (*Bubalus bubalis*, 2n = 50) cytogenetic map: assignment of 64 loci by fluorescence in situ hybridization and R-banding. *Cytogenetics and Genome Research* **102**, 65–75.
- Kierstein G., Vallinoto M., Silva A., Schneider M.P., Iannuzzi L. & Brenig B. (2004) Analysis of mitochondrial D-loop region casts new light on domestic water buffalo (*Bubalus bubalis*) phylogeny. *Molecular Phylogenetics and Evolution* **30**, 308–24.
- Kikkawa Y., Yonekawa H., Suzuki H. & Amano T. (1997) Analysis of genetic diversity of domestic water buffaloes and anoa based on variations in the mitochondrial gene for *cytochrome b*. *Animal Genetics* **28**, 195–201.
- Kumar S., Tamura K. & Nei M. (2004) MEGA3.1: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150–63.
- Kumar S., Gupta J., Kumar N., Dikshit K., Navani N., Jain P. & Nagarajan M. (2006) Genetic variation and relationships among eight Indian riverine buffalo breeds. *Molecular Ecology* **15**, 593–600.
- Lau C.H., Drinkwater R.D., Yusoff K., Tan S.G., Hetzel D.J.S. & Barker J.S.F. (1998) Genetic diversity of Asian water buffalo (*Bubalus bubalis*): mitochondrial DNA D-loop and *cytochrome b* sequence variation. *Animal Genetics* **29**, 253–64.
- Loftus R.T., MacHugh D.E., Bradley D.G., Sharp P.M. & Cunningham P. (1994) Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 2757–61.
- Meyer A., Kocher T.D., Basasibwaki P. & Willson A.C. (1990) Monophyletic origin of lake Victoria cichlid fish suggested mitochondrial DNA sequences. *Nature* **347**, 550–3.
- Moore S.S., Evans D., Byrne K., Barker J.S., Tan S.G., Vankan D. & Hetzel D.J. (1995) A set of polymorphic DNA microsatellites useful in swamp and river buffalo (*Bubalus bubalis*). *Animal Genetics* **26**, 355–9.
- Navani N., Jain P.K., Gupta S., Sisodia B.S. & Kumar S. (2002) A set of cattle microsatellite DNA markers for genome analysis of riverine buffalo (*Bubalus bubalis*). *Animal Genetics* **33**, 149–54.
- Nei M. (1976) Mathematical models of speciation and genetic distance. In: *Population Genetics and Ecology* (Ed. by S. Karlin, E. Nevo), pp. 723–66. Academic Press, New York, NY.
- Nei M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY.
- Nielsen R. & Wakeley J. (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**, 885–96.
- Ronquist F. & Huelsenbeck J.P. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–4.
- Sambrook J. & Russell D.W. (2001) Preparation and analysis of eukaryotic genomic DNA. In: *Molecular Cloning: A Laboratory Manual* (Ed. by J. Sambrook & D.W. Russell), pp. 6.11–6.14. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Shapiro B., Drummond A.J., Rambaut A. *et al.* (2004) Rise and fall of the beringian steppe bison. *Science* **306**, 1561–5.
- Steeves T.E., Anderson D.J. & Friesen V.L. (2005) The Isthmus of Panama: a major physical barrier to gene flow in a highly mobile pantropical seabird. *Journal of Evolutionary Biology* **18**, 1000–8.
- Tanaka K., Yamagata T., Masangkay J.S., Faruque M.O., Dang V.B., Salundik, Mansjoer S.S., Kawamoto Y. & Namikawa T. (1995) Nucleotide diversity of mitochondrial DNAs between the swamp and the river types of domestic water buffaloes, *Bubalus bubalis*, based on restriction endonuclease cleavage patterns. *Biochemical Genetics* **33**, 137–48.
- Tanaka K., Solis C.D., Masangkay J.S., Maeda K., Kawamoto I.Y. & Namikawa T. (1996) Phylogenetic relationship among all living species of the genus *Bubalus* based on DNA sequences of the *cytochrome b* gene. *Biochemical Genetics* **34**, 443–52.
- Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F. & Higgins D.G. (1997) The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**, 4876–82.

Supplementary Material

The following supplementary material is available for this article online from <http://www.blackwell-synergy.com/doi/full/10.1111/j.1365-2052.2007.01602.x>

Table S1 Alignment of mitochondrial *cytochrome b* gene sequences of the Indian river (present study) and swamp (Kikkawa *et al.* 1997) buffalo.

Figure S1 Maximum parsimony tree of river and swamp (Carabao) buffalo constructed from 945 bp sequences of mtDNA D-loop, with *Bos taurus* as an outgroup. Numbers above the branches correspond to the bootstrap using 1000 sequence replicates.

Figure S2 Reduced median network constructed using NETWORK 4.1.1.2 programme with 921 bp sequences of mtDNA D-loop region of the Indian river, Mediterranean

and swamp (Carabao) buffalo. The length of the line represents the number of mutational steps.

Figure S3 Reduced median network constructed using NETWORK 4.1.1.2 programme with 1125 bp sequences of 11 cytochrome b haplotypes of the Indian river and swamp

buffalo. The length of the line represents the number of mutational steps; red circles are median vectors.

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors.