Animal Genetics, 1998, **29**, 253–264

Genetic diversity of Asian water buffalo (*Bubalus bubalis*): mitochondrial DNA D-loop and cytochrome b sequence variation

C H Lau, R D Drinkwater, K Yusoff, S G Tan, D J S Hetzel, J S F Barker

Summary

Swamp and river buffalo mitochondrial DNA (mtDNA) was sequenced for 303 bp of the cytochrome b gene for 54 animals from 14 populations, and for 158 bp of the D-loop region for 80 animals from 11 populations. Only one cytochrome b haplotype was found in river buffalo. Of the four haplotypes identified in swamp buffalo, one found in all populations is apparently ancestral both to the other swamp haplotypes and to the river haplotype. The phylogenetic relationships among the 33 D-loop haplotypes, with a cluster of 11 found in swamp buffalo only, also support the evolution of domesticated swamp and river buffalo from an ancestral swamp-like animal, most likely represented today by the wild Asian buffalo (*Bubalus arnee*). The time of divergence of the swamp and river types, estimated from the D-loop data, is 28 000 to 87 000 years ago. We hypothesise that the species originated in mainland south-east Asia, and that it spread north to China and west to the Indian subcontinent, where the rive type evolved and was domesticated. Following domestication in China, the domesticated swamp buffalo spread through two separate routes, through Taiwan and the Philippines to the eastern islands of Borneo and Sulawesi, and south through mainland south-east Asia and then to the western islands of Indonesia.

Keywords: *Bubalus bulabis*, buffalo evolution, control region, cytochrome b, D-loop, genetic diversity, mitochondrial DNA

C H Lau
S G Tan
Department of Biology, Universiti Putra Malaysia, 43400 UPM, Serdang, Malaysia
R D Drinkwater
D J S Hetzel
CSIRO Division of Tropical Animal Production, Molecular Animal Genetics Centre, University of Queensland, St Lucia, Queensland 4072, Australia
K Yusoff
Department of Biochemistry and Microbiology, Universiti Putra Malaysia, 43400 UPM, Serdang, Malaysia
J S F Barker
Department of Animal Science, University of New England, Armidale NSW 2351, Australia

Introduction

The water buffalo of Asia (*Bubalus bubalis*) has been classified on morphological and behavioural criteria into two types (Macgregor 1939): the river buffalo of the Indian subcontinent and west to the Balkans and Italy, and the swamp buffalo of south-east Asia, from Assam and Nepal east to the Yangtze valley of China. These types differ in chromosome number (Ulbrich & Fischer 1967; Fischer & Ulbrich 1968): swamp, $2n = 48$, river, $2n = 50$, and for allele frequencies at protein-coding loci (Amano 1983; Barker *et al*. 1997a) and at microsatellite loci (Barker *et al*. 1997b).

Previous studies of mitochondrial DNA (mtDNA) variation using restriction fragment length polymorphism (Amano *et al*. 1994; Tanaka *et al*. 1995) have shown genetic differences between the two types, and have considered their origin, divergence and domestication. We present here results for DNA sequence variation in mtDNA control region (D-loop) and cytochrome b. The D-loop is the most variable portion of the mammalian mtDNA genome, and is commonly variable at the intraspecific level, making it useful for studies of genetic variability among populations and phylogenetic analysis, while cytochrome b usually has only moderate levels of intraspecific variation. Thus our aims are to compare D-loop variation, particularly among swamp type populations, with our previous studies of protein-coding and microsatellite loci (Barker *et al*. 1997a,b), and to compare the swamp and river types for cytochrome b variation.

Materials and methods

Sample collection

Samples from 80 animals were used to obtain mtDNA D-loop sequences, and from 54 animals for cytochrome b sequences, with 30 animals common to both sets (Table 1). Animals from populations in Thailand other than Surin were included for cytochrome b analysis after three different haplotypes were found in four animals from Surin (Table 1). For each population, the animals assayed were a subset of those used by Barker *et al*. (1997a), see their Fig. 1 for the location of each population. Although Lankan buffalo are phenotypically swamp type, previous studies (Barker *et al*. 1997a,b; and references therein) have shown them to be clearly

Correspondence: J S F Barker.

Accepted 14 March 1998

Table 1. Populations and numbers of animals assayed for D-loop and cytochrome b sequence variation, and frequencies of each of the cytochrome b haplotypes

Population	No. of animals		Cyt b haplotype frequencies				
	D-loop	Cyt b	Swamp 1	Swamp 2	Swamp 3	Swamp 4	Murrah 1
Thailand							
Chiang Mai	–	7	2	4		1	
Kam Paeng Seng	–	2	2				
Surin	7	4	1	1	2		
Hat Yai	–	1	1				
Malaysia							
Trengganu	7	5	3		2		
Sabah	8	3	3				
Sarawak	8	4	4				
Philippines	7	4	4				
Indonesia							
Bogor	7	4	4				
Sulawesi	9	4	4				
Australia	4	4	4				
Sri Lanka South	8	3					3
Murrah							
Sri Lanka	5	5					5
Malaysia	10	4					4
Total	80	54	32	5	4	1	12

river type, and they are included with the two Murrah populations as river buffalo in all analyses here.

DNA extraction

Blood collection and treatment was as described by Barker *et al*. (1997a). DNA was prepared from buffy coat cell mixtures that were obtained by centrifugation of fresh whole blood. Buffy coat cells were removed from the blood sample by aspiration, and stored at –70 °C. For DNA extraction, 25 μl of buffy coat cells were mixed with 0·5 ml of 1× TE (10 mM Tris–HCl pH 8·0, 1 mM EDTA) by brief vortexing and spun for 30 s on a bench microfuge at room temperature. The pellet was resuspended in 0·5 ml TE and spun as before. After discarding the supernatant, the pellet was dissolved in 100 μl proteinase K solution (0·1 mg/ml proteinase K in 10 mM EDTA) and incubated at 56 °C for 45 min, followed by an incubation at 95 °C for 10 min to deactivate the protease. Each sample was extracted with a phenol/chloroform/iso amyl alcohol mix, precipitated in ethanol, and resuspended in 1× TE. The DNA solution was stored at 4 °C until used.

Polymerase chain reaction and sequencing: D-loop

Two compatible oligonucleotide primers were designed from the bovine mitochondrial D-loop sequence, viz. Bov A (5′-GAC CTC TAT AGC AGT ACA TA-3′) and Bov B (5′-CAT TAG TCC ATC GAG ATG TC-3′). These primers were positioned at 16 047–16 066 bp and 16 326–00007 bp, respectively, in the bovine mitochondrial sequence (Anderson *et al*. 1982), and were used in a PCR to generate a fragment of the mitochondrial DNA from a sample of Murrah buffalo DNA. This fragment was cloned into a pUC-18 vector and sequenced. From this sequence, two buffalo specific oligonucleotide primers were derived for all subsequent PCR and sequencing analysis on buffalo DNA, viz. Buff2A (5′-CAT GCA TGA TAG TAC ATA GTA-3′) and Buff2B (5′-GAG ATG GCC CTG AAG AAA GAA C-3′). Primers positioned within the Buff2A–Buff2B sequence were used for all sequencing analysis of the buffalo mitochondrial derived PCR fragments. The sequencing primers were: Buff3A (5′-GTA CAT AGC ACA TTT AAG AC-3′) and Buff2B.

The mixes for PCR were: 60 ng DNA, 150 μM of each dNTP, 2·0 mM $MgCl_2$, 1× reaction buffer (50 mM KCl, 10 mM Tris–HCl, pH 8·3), 100 ng of each primer, 5 units of *taq* polymerase (Perkin Elmer, Branchburg, NJ) in a reaction volume of 50 μl. PCR thermal profiles consisted of a single 2 min cycle at 95 °C, followed by 35 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 45 s, and concluded with a single cycle of 72 °C for 5 min. A PHC-2-Dri-Block cycler connected to a CH5/PC chiller and pump (Techne, Duxford, Cambridge, UK) was used for temperature

cycling. The PCR products from each sample were purified prior to sequencing analysis in a CHROMA SPIN-100 chromatography column (CLONTECH, Palo Alto, CA) according to the manufacturer's instructions, and the sample recovered in 50 μl of 0·1× TE.

The purified samples generated from the PCR were sequenced using the CircumVent Thermal Cycle Dideoxy Sequencing kit (New England Biolabs, Beverly, MA) with $Vent_R$(exo-) DNA polymerase, with sequencing reactions initiated with 1·2 pmol of Buff2B primer that had been end-labelled with [γ-^{33}P]ATP (2000 Ci/mmol, 10 mCi/ml) (NEN DuPont, Boston, MA). The thermal profile used was: 95 °C for 1 min, followed by 25 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 45 s. The sequencing products were electrophoresed in 45 cm × 30 cm × 0·4 mm 8% (w/v) denaturing (8 M urea) polyacrylamide gels in 1 × TBE, then the gels were dried and autoradiographed. Analysis of the sequence ladders was carried out manually by at least three independent reads. For 34 of the sequences, one or more nucleotide positions showed double bands with similar intensity which could not be clearly differentiated. These double bands were observed for A and G, C and T, and A and T, and these bases were designated as '?' (unknown) for sequence analysis.

Polymerase chain reaction and sequencing: cytochrome b

Amplification of the cytochrome b fragment and sequencing were done as described by Lau *et al.* (1995).

DNA sequences of the D-loop and of cytochrome b were separately aligned with the bovine sequence (Anderson *et al.* 1982), using the default options of CLUSTAL W 1·4 (Thompson *et al.* 1994). The 158 bp D-loop sequence corresponds to nucleotides 16 089–16 245 of the bovine mtDNA sequence, while the 303 bp cytochrome b sequence corresponds to nucleotides 14 613–14 915 (position numbers as in Anderson *et al.* 1982). GenBank accession numbers for the cytochrome b sequences Swamp 2 and Murrah 1 are X78960 and X84906, respectively, and for the D-loop Aust 1 sequence is AF016397.

Statistical analyses

In all analyses of D-loop haplotype differences, pair-wise deletion was used to allow for missing bases. Thus the number of nucleotide differences between haplotypes, for any pair where bases are missing, are minimum estimates. Intraspecific variation was summarised by computing haplotype diversity (h) and nucleotide diversity (π) (Nei & Tajima 1981; Nei 1987) using the REAP computer package (McElroy *et al.* 1991). Relationships among D-loop haplotypes were inferred in two ways. First, as the mean Jukes–Cantor estimate of the number of nucleotide substitutions per site between the different sequences was 0·06, and the transition/transversion ratio was 6·0, we followed the guidelines of Kumar *et al.* (1993) and used the Kimura two-parameter model (Kimura 1980) to estimate nucleotide divergence between all pairs of haplotypes. These then were used to generate a neighbour-joining tree (Saitou & Nei 1987), as implemented in the MEGA computer package (Kumar *et al.* 1993). Second, a minimum spanning tree (Prim 1957), based on the number of nucleotide differences between each pair of haplotypes, was produced using the MST program in the NTSYS package (Rohlf 1988). Nucleotide divergence between populations (d_A; Nei 1987; Eqn. 10·21) was estimated using REAP, and a neighbour-joining tree then generated using MEGA. Nucleotide divergence (d_A) was estimated by $d_A = d_{XY} - (d_X + d_Y)/2$,where d_{XY} is the average number of nucleotide substitutions between DNA haplotypes from populations X and Y, and d_X and d_Y are the average numbers of nucleotide substitutions for a randomly chosen pair of haplotypes in each population. However, Nei (1985) indicated that the pattern of population differentiation may be inferred using d_{XY}, which has a smaller coefficient of variation than d_A. Thus d_{XY} estimates (also from REAP) were used in MEGA to generate a neighbour-joining tree. Population differentiation also was assessed using (i) the exact test (Raymond & Rousset 1995a) in the GENEPOP package (Raymond & Rousset 1995b, 1997), with significance levels determined by applying to the probability levels calculated by GENEPOP the sequential Bonferroni procedure (Hochberg 1988; Lessios 1992), and (ii) ϕ_{ST} estimated from all pair-wise population comparisons by AMOVA (Excoffier *et al.* 1992; Excoffier 1995).

For cytochrome b, nucleotide divergence between pairs of haplotypes estimated as the proportion of nucleotide differences (*p* distance), Jukes–Cantor distance or Kimura two-parameter distance were all essentially the same. Thus a neighbour-joining tree was generated using only the *p* distances.

Estimates of nucleotide divergence between populations (d_A) and assumed rates of nucleotide substitution per site per year (λ) for the D-loop region were used to estimate the time of

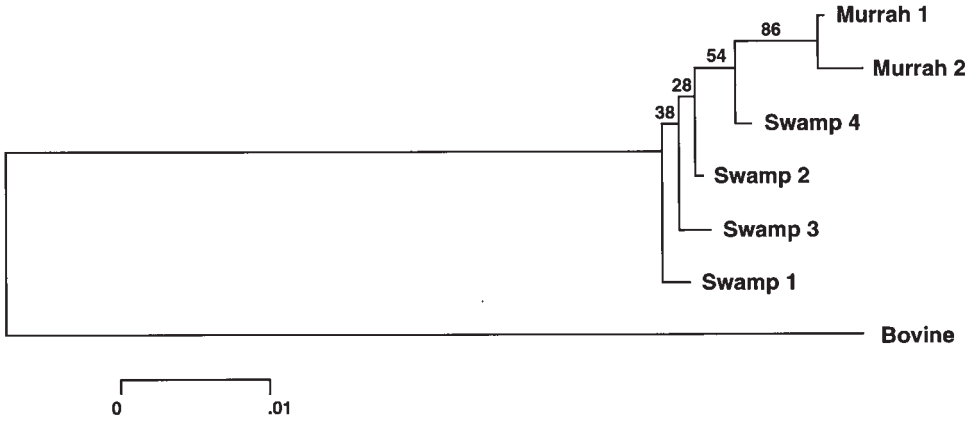

Fig. 1. Neighbour-joining tree for six cytochrome b haplotypes of water buffalo, with the bovine sequence as an outgroup. Numbers on the nodes are percentage bootstrap values from 1000 replications, and a scale bar for branch lengths is shown.

divergence (t, in years) between the swamp and river buffalo types, and among some of the swamp populations, where $d_A = 2\lambda t$ (Nei 1987; Eqn. 10·22).

Results

Cytochrome b

Our sequencing of a 303-bp portion of cytochrome b revealed five haplotypes and five polymorphic nucleotide positions (Table 2). An additional haplotype (designated Murrah 2 in Table 1) was found in a cytochrome b sequence obtained from GenBank (accession no. D34638; Kikkawa *et al.* unpublished). This haplotype includes a transversion substitution at nucleotide position 14 909, but the other five variable positions are transition substitutions. All of the nucleotide substitutions were silent, with no changes in amino acid composition.

The haplotype Murrah 1 was found only in river buffalo (Murrah and Lankan), and all other haplotypes in swamp buffalo only. River and swamp buffaloes differed diagnostically at two nucleotide positions (14 726 and 14 867; Table 2). In swamp buffalo, all four haplotypes were found in the mainland Asian populations (Thailand and Peninsular Malyasia), but only the most common of these (Swamp 1) was present in the island populations of south-east Asia and in Australia.

The neighbour-joining tree of the six haplotypes (Fig. 1) shows a clear separation of those

Table 2. Polymorphic nucleotide sites for six cytochrome b haplotypes observed in 54 swamp and river buffalo, and for the bovine sequence (sites variable only for buffalo vs. bovine indicated by * below the bovine sequence)

	Nucleotide position											
	111	111	111	111	111	111	111	111	111	111	111	11
	444	444	444	444	444	444	444	444	444	444	444	44
	666	666	677	777	777	777	888	888	888	888	888	99
	112	234	501	112	344	588	012	233	355	677	777	01
	584	792	481	276	817	006	172	814	758	701	347	92
SWAMP 1	CTA	CCG	CCC	GCG	TAT	TAC	AAC	AAT	CCC	CAC	ATG	AG
SWAMP 2	...	.T.	...	...	...	...	...	...	...	...	...	..
SWAMP 3	...	.T.	.T.	...	...	...	...	...	...	...	...	..
SWAMP 4	...	.T.	...	...	...	...	...	...	...	..T	...	..
MURRAH 1	...	.T.	...	..A	...	...	...	...	...	T.T	...	..
MURRAH 2	...	.T.	...	..A	...	...	...	...	...	T.T	...	T.
BOVINE	TCG	A.A	ATT	ATA	CCC	CGT	GTT	GTC	TTT	.T.	GCA	CA
	***	* *	* *	**	***	***	***	***	***	*	***	*

The nucleotide positions are those for the bovine sequence (Anderson *et al.* 1982)

Table 3. Polymorphic nucleotide sites for 33 D-loop haplotypes observed in 80 swamp and river buffalo, and for the bovine sequence (sites variable only for buffalo vs. bovine indicated by * below the bovine sequence)

	Nucleotide position					
	11111	11 1111111	1111111111	1111111111	1111111111	1111111111
	66666	66 6666666	6666666666	6666666666	6666666666	6666666666
	00000	00 1111111	1111111111	1111111111	1111111111	1222222222
	89999	99 0011111	1112222233	3333334444	6666688999	9001223334
	90123	45 1912346	7891256701	2345671237	2345626234	7049890121
Swamp and river						
Aust 1	CATAG	CATACATTCT	ATCACGCG--	------CTTC	GATCACCTTC	GATCGTTATA
Aust 2	.....	C	....	.........T	.....T....	.G....C...
Swamp only						
Aust 3	..A..	.GC..G.C..	G......A..	.......C.T	..C.......	AGC.ACCT..
Aust 4	.....	.......?.?	.?........	.......C..	....	......C...
Phil 1	.....	?GC..G...C	....?..?..	.......?.?	.?...?...?	?G....C?..
Phil 2	.....	TGC..G.C.C	GC.....A..	.......C..	.G........	AG....CT..
Phil 3	.....	.........?	....	....	....	.?....C...
Phil 4	....	T.........	.......?..	....	......G...	?.....?...
Phil 5	.....	.........?	.......A..	.........?	.....?....	.G....?...
Phil 6	.....	.........?	G......A..	....	....	.G....C...
Phil 7	.....	T....G.?.C	GC.....?..	.......C..	.G........	.G....C...
Bogor 1	...CA	T....?...C	?C........	.......C..	.?........	.G....C...
Sula 1	.....	.........G	G.........	....	....	......?...
Sula 2	...C.	TGC..G.C.C	.?.....A..	.......C..	.?........	AG....C?..
Sula 3	.....	.GC..G.C..	G......A..	.......C.T	..C.......	AGC.ACCT..
Thai 1	.....	.?...?....	........?..	.......?.?	.....T....	?G....C...
Thai 2	.....	TGC..G.C..	GC..T..A..	.........T	..C.......	AGC.A.CT..
Thai 3	TCG..	..C..G.C..	G......A..	.......C.T	..C......?	AGC.ACCT..
Thai 4	.....	C	....	......T.C.	.....?....	.G....C...
Thai 5	.....	.G....CC.C	G......A..	.........T	....	AGC.AC.T..
Thai 6	.....	C	....	........C?	.?...?....	?G....C...
Sabah 1	.....	?........C	G......A..	....	....	.G........
Sara 1	T....	TGC..G.C.C	GC.....A..	.......C..	.G........	AG....CT..
Sara 2	.....	T-.......C	....	.........T	....TT....	.G....C...
Sara 3	.....	C	....	.........T	....	.G....CT..
River only						
SriL 1	.....	....	.C........	....	...T.T....	.G...CC...
SriL 2	.....	....	....	....	....	.....CC...
SriL M1	.....	C	....	....	.....T....	.G....C...
SriL M2	.....	C	....	.........T	.....T....	.G.T..C...
SriL M3	.....	....	....	.........T	.....T....	.G....C...
Mal M1	.....	C	....	.........T	.....T....	.G........
Mal M2	.....	C	....	....	....	....
Mal M3	.....	........TC	....	.........T	.....T....	.G....C...
BOVINE	.....	T.-TT.....	GATGTATCTT	ATATATT.AT	A..T...GCT	.G..AAACCC
		**	** ** **	******	* **	**

Nucleotides labelled ? were scored as missing because of double bands of similar intensity. The nucleotide positions are those for the bovine sequence (Anderson *et al.* 1982)

found in river and in swamp buffalo, but no strong support for the pattern of relationships among those found only in swamp buffalo.

D-loop

Thirty-three D-loop haplotypes were observed (Table 3), and including the bovine sequence, there are 55 variable sites in the total of 158 nucleotides. In the buffalo, 36 sites were polymorphic, while 19 additional sites (including the buffalo deletion at 16 130–16 137) differed between buffalo and bovine. The single inserted nucleotide in the buffalo sequences is between positions 16 099 and 16 100 in the bovine sequence. Because of missing nucleotides at site 16 194, this site also may be variable only for bovine vs. buffalo, and not polymorphic in buffalo. Also because of missing nucleotides, either Thai 1 and Phil 5, or Thai 1

Table 4. Frequencies of 33 D-loop mitochondrial haplotypes in 11 populations of water buffalo

	Population											
	Swamp buffalo								River buffalo			
Haplotype	Surin (7)*	Trengganu (7)	Sabah (8)	Sarawak (8)	Philippines (7)	Bogor (7)	Sulawesi (9)	Australia (4)	Sri Lanka South (8)	Murrah Sri Lanka (5)	Murrah Malaysia (10)	Overall frequency
Aust 1	–	–	–	–	–	0·714	–	0·250	0·625	0·200	–	0·1500
Aust 2	0·143	–	–	0·125	–	–	0·111	0·250	0·125	0·200	0·700	0·1625
Aust 3	–	–	–	–	–	–	–	0·250	–	–	–	0·0125
Aust 4	–	–	–	–	–	–	–	0·250	–	–	–	0·0125
Phil 1	–	–	–	–	0·143	–	–	–	–	–	–	0·0125
Phil 2	–	0·143	0·250	0·500	0·143	–	0·556	–	–	–	–	0·1625
Phil 3	–	0·714	0·125	–	0·143	0·143	–	–	–	–	–	0·1000
Phil 4	–	–	–	–	0·143	–	–	–	–	–	–	0·0125
Phil 5	–	–	–	–	0·143	–	–	–	–	–	–	0·0125
Phil 6	–	–	0·500	–	0·143	–	–	–	–	–	–	0·0625
Phil 7	–	–	–	–	0·143	–	–	–	–	–	–	0·0125
Bogor 1	–	–	–	–	–	0·143	–	–	–	–	–	0·0125
Sula 1	–	–	–	–	–	–	0·111	–	–	–	–	0·0125
Sula 2	–	–	–	–	–	–	0·111	–	–	–	–	0·0125
Sula 3	–	0·143	–	–	–	–	0·111	–	–	–	–	0·0250
Thai 1	0·143	–	–	–	–	–	–	–	–	–	–	0·0125
Thai 2	0·143	–	–	–	–	–	–	–	–	–	–	0·0125
Thai 3	0·143	–	–	–	–	–	–	–	–	–	–	0·0125
Thai 4	0·143	–	–	–	–	–	–	–	–	–	–	0·0125
Thai 5	0·143	–	–	–	–	–	–	–	–	–	–	0·0125
Thai 6	0·143	–	–	–	–	–	–	–	–	–	–	0·0125
Sabah 1	–	–	0·125	–	–	–	–	–	–	–	–	0·0125
Sara 1	–	–	–	0·125	–	–	–	–	–	–	–	0·0125
Sara 2	–	–	–	0·125	–	–	–	–	–	–	–	0·0125
Sara 3	–	–	–	0·125	–	–	–	–	–	–	–	0·0125
SriL 1	–	–	–	–	–	–	–	–	0·125	–	–	0·0125
SriL 2	–	–	–	–	–	–	–	–	0·125	–	–	0·0125
SriL M1	–	–	–	–	–	–	–	–	–	0·200	–	0·0125
SriL M2	–	–	–	–	–	–	–	–	–	0·200	–	0·0125
SriL M3	–	–	–	–	–	–	–	–	–	0·200	–	0·0125
Mal M1	–	–	–	–	–	–	–	–	–	–	0·100	0·0125
Mal M2	–	–	–	–	–	–	–	–	–	–	0·100	0·0125
Mal M3	–	–	–	–	–	–	–	–	–	–	0·100	0·0125

*Number of animals sampled from each population.

and SriL M1 could be the same haplotype. For each of sites 16 091 and 16 116, two mutations have occurred in the water buffalo, as three different nucleotides were found. The frequencies of each haplotype in each population and overall are given in Table 4. A striking feature of these frequencies is the high proportion (27/33) that were found in only one population.

Haplotype and nucleotide diversities are given in Table 5. Two of the three most common haplotypes (Aust 1 and Aust 2) were found in both swamp and river buffaloes, 23 haplotypes were found in swamp only, and eight in river only, but haplotype diversities are not significantly different ($t = 1{\cdot}636$, $P < 0{\cdot}2$) between swamp and river buffalo. Within population haplotype diversities were highest for Surin, Philippines, Australia and Sri Lanka Murrah, where for each of these populations, each animal sampled had a different haplotype. Overall haplotype and nucleotide diversities were 0·917 and 0·053.

The relationships among the 33 haplotypes (Fig. 2) show two major clusters, one of 11 found in swamp buffalo only, and the remaining 22 that include haplotypes found in swamp only, river only, and in both swamp and river buffalo. Although the bootstrap support for separation of these clusters is low, the same separation and essentially the same tree was obtained using the *p*-distance and Jukes–Cantor distance with MEGA, and a similar separation using the DNADIST program in PHYLIP (Felsenstein 1993). The genealogical mixing of haplotypes from both swamp and river buffalo in the second cluster must reflect polymorph-

Table 5. Haploytpe diversity (h) and nucleotide diversity (π) estimated from mtDNA D-loop sequence for each population of swamp and river buffalo

Buffalo type/ population	Number of haplo-types	Haplotype diversity ($h \pm$ SD)	Nucleotide diversity
Swamp	25	0·9135 ± 0·0224	0·0561
Surin	7	1·0000 ± 0·0764	0·0749
Trengganu	3	0·5238 ± 0·2086	0·0487
Sabah	4	0·7500 ± 0·1391	0·0355
Sarawak	5	0·7857 ± 0·1508	0·0545
Philippines	7	1·0000 ± 0·0764	0·0380
Bogor	3	0·5238 ± 0·2086	0·0169
Sulawesi	5	0·7222 ± 0·1592	0·0518
Australia	4	1·0000 ± 0·1768	0·0714
River	10	0·7984 ± 0·0667	0·0203
Sri Lanka	4	0·6429 ± 0·1841	0·0195
Murrah Sri Lanka	5	1·0000 ± 0·1265	0·0192
Murrah Malaysia	4	0·5333 ± 0·1801	0·0079
Overall	33	0·9174 ± 0·0157	0·0526

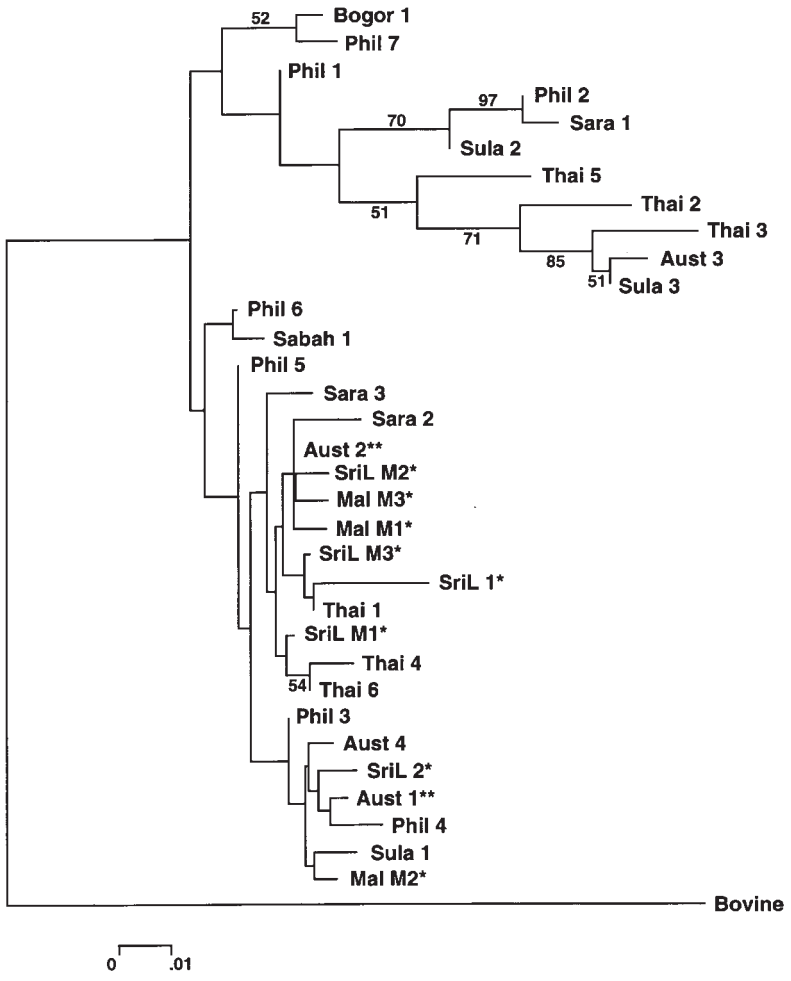

Fig. 2. Neighbour-joining tree for 33 mtDNA D-loop haplotypes of water buffalo, with the bovine sequence as an outgroup. Numbers on the nodes are percentage bootstrap values from 1000 replications (only those greater than 50% are shown), and a scale bar for branch lengths is shown. **Haplotypes found in both swamp and river buffalo; *haplotypes found in river buffalo only.

ism in the ancestral population (Nei 1987), but the generally low bootstrap values preclude interpretation of the haplotype relationships.

However, the minimum spanning tree (Fig. 3) gives these evolutionary relationships, in showing putative ancestral-descendant patterns. The cluster of haplotypes found only in swamp buffalo (diverging from Phil 1), is closest to the bovine, and is clearly separated (cluster A), while the remaining haplotypes cluster on Phil 5 (cluster B) and Phil 3 (cluster C). Haplotypes that were found in more than one population occur in each cluster, namely Sula 3 and Phil 2 in cluster A, Aust 2 and Phil 6 in cluster B, and Aust 1 and Phil 3 in cluster C. Although these shared haplotypes generally do not occur only in geographically close populations, Phil 2 (cluster A), Phil 6 (cluster B) and Phil 3 (cluster C) were found predominantly in the eastern island populations of the Philippines, Sabah, Sarawak and Sulawesi.

Estimates of nucleotide divergence between populations (Table 6) and the neighbour-joining tree derived from them (Fig. 4) show the populations grouped into two clusters. Although the relationships among the populations show some congruity with their geography, Bogor is clearly exceptional in clustering with the three river buffalo populations. In contrast, while the neighbour-joining tree derived from d_{XY} divergences (Fig. 5) still shows Bogor clustered with the river buffalo populations, the pattern of relationships among the other swamp buffalo populations is more congruent with their geography, and the known origin of the Australian population from Indonesia (Barker *et al.* 1997a,b). In pair-wise tests of population differentiation (Table 7), 30 of 55 had *P*-values less than 0·05, but only three of these were significant after applying the Bonferroni procedure. However, estimates of ϕ_{ST} (Table 7) show more extensive population differentiation, particularly between the swamp and river populations.

AMOVA analyses showed significant differences between the swamp and river types, and among populations within the swamp type, but not the river type. For the regional grouping of populations, based on type and then within the swamp type, combining geography and population differentiation, the variance among populations within regions was not significant. The between regions variance, although significant, did not indicate very strong structuring.

Estimation of divergence times

The average d_A distance between swamp and

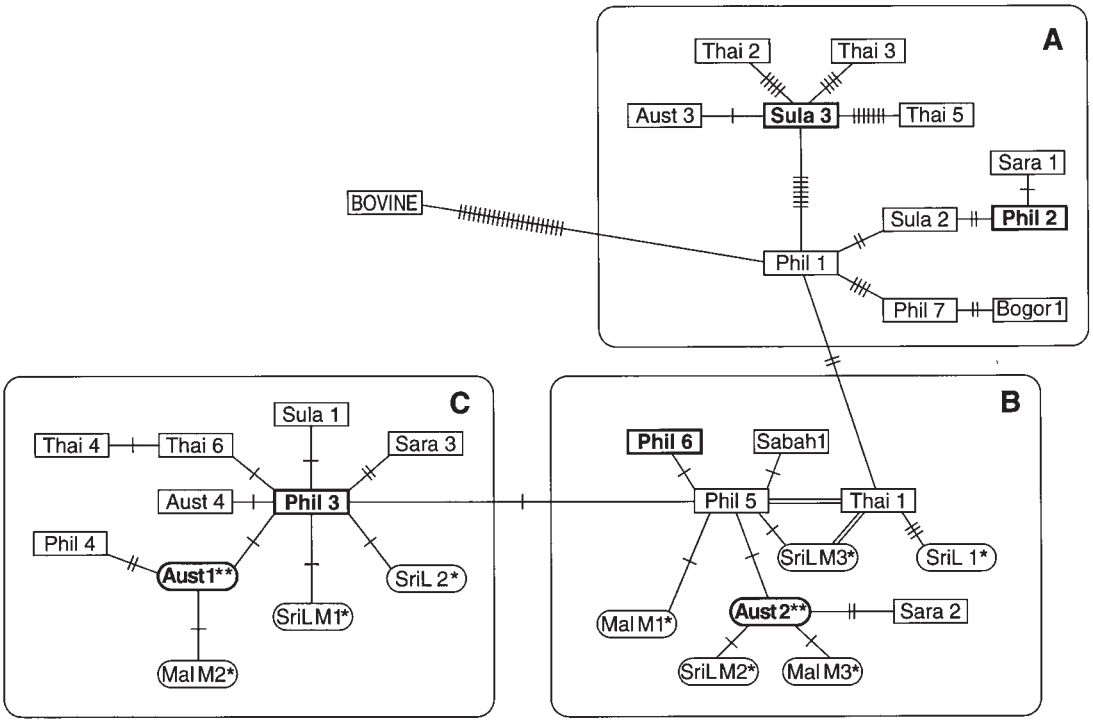

Fig. 3. Minimum spanning tree showing the network of interrelationships among the 33 mtDNA D-loop haplotypes. Capital letters (A, B and C) identify the clusters referred to in the text. Haplotypes in bold are those found in more than one population. **Haplotypes found in both swamp and river buffalo; *haplotypes found in river buffalo only. Ticks indicate the inferred number of mutational steps between pairs of haplotypes.

river buffalo populations (Table 6) is 0·0206. Using an average evolutionary rate estimated from human D-loop data of $0{\cdot}118 \times 10^{-6}$ substitutions per site per year (Stoneking *et al.* 1992), the divergence time estimate is ≈87 000 bp. However, this assumes that evolutionary rates are similar in humans and water buffalo. As nuclear DNA apparently evolves faster in cattle than in humans (Li *et al.* 1990; Bulmer *et al.* 1991), so too might mtDNA evolve faster in cattle and in other species of the tribe Bovini (including *Bubalus bubalis*, the water buffalo). Using the average sequence divergence between African and European breeds of cattle of 0·0073 (Loftus *et al.* 1994), and assuming this has accumulated largely since domestication (≈10 000 years ago), the evolutionary rate estimate for cattle is $0{\cdot}365 \times 10^{-6}$ substitutions per site per year (i.e. about three times the rate for humans). On this basis, the swamp-river divergence time is ≈28 000 bp. As a second approach to scaling time and evolutionary rates, we have used the time of domestication of the swamp buffalo in China at about 7000 bp (Chen & Li 1989), and its likely separate spread south through mainland Asia and through the Philippines (Barker *et al.* 1997b).

Table 6. Matrix of d_A distances (D-loop nucleotide divergence between populations) among 11 water buffalo populations

	Swamp							River		
	Trengganu	Sabah	Sarawak	Philippines	Bogor	Sulawesi	Australia	Sri Lanka	Murrah Sri Lanka	Murrah Malaysia
Surin	–0·0022	0·0055	0·0097	0·0031	0·0214	0·0133	–0·0078	0·0191	0·0109	0·0138
Trengganu		0·0005	0·0101	–0·0041	0·0053	0·0137	–0·0104	0·0055	0·0063	0·0132
Sabah			0·0083	–0·0030	0·0174	0·0115	0·0021	0·0202	0·0188	0·0248
Sarawak				0·0035	0·0418	–0·0038	0·0125	0·0462	0·0375	0·0397
Philippines					0·0109	0·0083	–0·0035	0·0134	0·0115	0·0172
Bogor						0·0486	0·0017	–0·0007	0·0100	0·0196
Sulawesi							0·0158	0·0538	0·0499	0·0540
Australia								0·0006	0·0010	0·0070
Sri Lanka									0·0064	0·0157
Murrah Sri Lanka										–0·0001

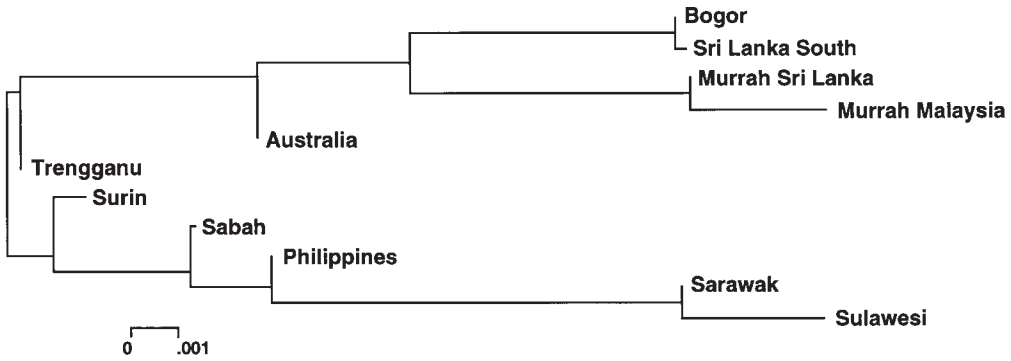

Fig. 4. Neighbour-joining tree for 11 water buffalo populations, based on nucleotide divergence (d_A) between populations, with negative divergence estimates set to zero. A scale bar for branch lengths is shown.

Assuming the divergence between the Surin and Philippines population (0·0031; Table 6) has accumulated since this split to different dispersal routes, using the human rate gives an estimate of this separation of ≈13 000 bp. In contrast, the faster rate derived from the cattle data gives an estimate of ≈4000 bp, much more in accord with the presumed time of domestication and spread in conjunction with rice cultivation (Bellwood 1992).

Discussion

Cytochrome b haplotypes are distinct in the swamp and river buffalo, with four found in swamp only and two in river buffalo only (Fig. 1). The phylogenetic relationships of the four haplotypes found in swamp buffalo are not clearly defined (low bootstrap support for these nodes in Fig. 1). However, Swamp 1 is clearly ancestral, as it was found in all swamp buffalo populations, and it appears to be ancestral also to the river buffalo haplotypes. This phylogeny, together with the presence of all four swamp haplotypes in Thailand, suggests that the species may have originated in this area of mainland south-east Asia, and subsequently spread north to China, and west to the Indian subcontinent, where the river type evolved. This suggested evolution of the river type from an ancestral swamp-like animal is further supported by the morphological similarity of the swamp type to the wild Asian buffalo (*Bubalus arnee*), and by the finding that the phenotypically swamp type Lankan buffalo are genetically river type (Amano *et al.* 1982; Barker *et al.* 1997a,b), and thus represent the type of buffalo from which the present breeds of river buffalo of the Indian subcontinent were derived (Barker *et al.* 1997a). The phylogeny of the D-loop haplotypes (Figs 2 and 3) provides further strong support for this hypothesis. The neighbour-joining tree (Fig. 2) shows two major clusters, one of haplotypes found in swamp buffalo only (also seen as cluster A in Fig. 3), and one of haplotypes found in both swamp and river buffalo (the derived clusters B and C in Fig. 3). Most significantly, all haplotypes that were found only in river buffalo occur as terminal tips (i.e. most recent) in the minimum spanning tree (Fig. 3). Tanaka *et al.* (1996), using complete cytochrome b sequences, also

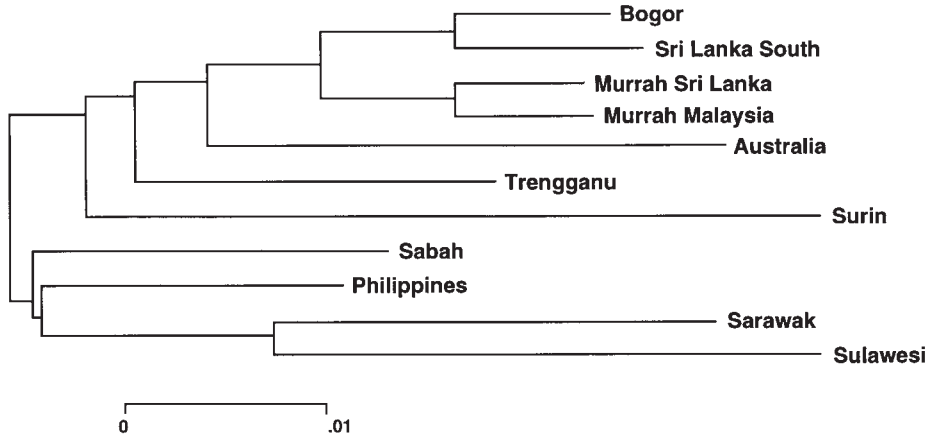

Fig. 5. Neighbour-joining tree for 11 water buffalo populations, based on average number of nucleotide substitutions between populations (d_{XY}). A scale bar for branch lengths is shown.

Table 7. Tests of population differentiation (i) above diagonal — significant *P*-values from exact tests using GENEPOP (*P*-values in italics indicate the only comparisons that were significant after applying the sequential Bonferroni procedure), and (ii) below diagonal — estimates of ϕ_{ST} obtained using AMOVA

	Swamp						River				
	Surin	Trengganu	Sabah	Sarawak	Philippines	Bogor	Sulawesi	Australia	Sri Lanka	Murrah Sri Lanka	Murrah Malaysia
Surin		0·022	0·038	–	–	0·020	0·030	–	0·025	–	0·038
Trengganu	–0·031		0·035	0·025	–	0·015	0·016	0·044	0·003	0·029	*0·001*
Sabah	0·107	0·023		–	–	0·003	0·044	–	0·002	0·048	*0·001*
Sarawak	0·139	0·166	0·162***		–	0·006	–	–	0·005	–	0·006
Philippines	0·054	–0·100	–0·086	0·070		0·021	–	–	0·025	–	0·006
Bogor	0·319***	0·135	0·388***	0·524***	0·282*		0·003	–	–	–	*0·001*
Sulawesi	0·185	0·213	0·206	–0·075	0·146	0·559***		–	0·003	–	0·004
Australia	–0·118	–0·189	0·098	0·190*	–0·029	0·121	0·233		–	–	–
Sri Lanka	0·301***	0·140*	0·414***	0·546***	0·316***	–0·042	0·583***	0·100		–	0·003
Murrah Sri Lanka	0·144	0·123	0·372*	0·455*	0·257*	0·363**	0·532***	0·052	0·245		–
Murrah Malaysia	0·307***	0·370***	0·554***	0·581***	0·471***	0·636***	0·649***	0·339***	0·548***	0·041	

$^{*}P < 0{\cdot}05$, $^{**}P < 0{\cdot}01$, $^{***}P < 0{\cdot}001$.

have shown a close phylogenetic relationship between the river type and the Sri Lankan wild buffalo, and have suggested that the wild buffaloes of the Indian subcontinent were the source of the river type. This agrees with our results, but we note that in our hypothesis, the wild Sri Lankan buffalo do not represent the ancestor of both the swamp and river types. The wild Asian buffalo (*Bubalus arnee*), which may still exist in Assam, Nepal and Indochina (Mason 1974), more likely is the ancestor of both the swamp and river types, but genetic data to support this hypothesis is lacking.

The time of divergence of the swamp and river types has been estimated as at least 10 000–15 000 years ago (Barker *et al.* 1997b), 28 000–87 000 years ago (this paper, D-loop divergence), 700 000 years or more (Tanaka *et al.* 1995), within the last 1 Myr (Amano *et al.* 1994), and 1.7 Myr (Tanaka *et al.* 1996). This variation in estimated divergence times is likely a function of the different rates of evolution of the genes and sequences used, while the longer estimates based on single genes may be overestimates, as the time of gene splitting may be much earlier than the time of population splitting (Takahata & Nei 1985). However, genetic divergence of the two buffalo types is not just a function of domestication, which has taken place much more recently. Buffalo had apparently been domesticated in India 5000 years ago and in China 4000 years ago (Cockrill 1974), although Chen & Li (1989) suggest domestication in China at least 7000 years ago.

The genetic relationships among the 11 populations, based on nucleotide divergence for the D-loop haplotypes (Fig. 4), do not accord well with their relationships derived from microsatellite or protein coding loci (Barker *et al.* 1997b). However, Barker *et al.* (1997b) noted that the relationships based on protein coding loci were distorted by bottleneck effects on some populations. The unexpected clustering here of the Bogor swamp population with the river buffalo populations is due primarily to the high frequency of the Aust 1 haplotype in the Bogor population (Table 4), and is probably a function of the small sample sizes studied. The swamp buffalo is thought to have spread from mainland Asia to the islands of south-east Asia following domestication in China (Bellwood 1992; Barker *et al.* 1997b), so that other differences between the mtDNA D-loop and microsatellite phylogenies likely reflect differential male and female migration, and/or large standard errors of the d_A distances (Takahata & Nei 1985). While the tree derived from d_{XY} distances (Fig. 5) is more congruent with the geography of the swamp populations, and is likely a better representation of their pattern of differentiation, we conclude that the microsatellite based tree gives a more accurate representation of the genetic relationships among these populations.

Nevertheless, the D-loop results do support the suggestion of Barker *et al.* (1997b) that the dispersal of domesticated swamp buffalo from mainland Asia followed two routes, from China south through Thailand and Malaysia, and from China through the Philippines to the eastern islands of Borneo (Sabah and Sarawak) and Sulawesi. Firstly, the haplotypes Phil 2, Phil 3 and Phil 6 are found predominantly in the eastern island populations. Secondly, the averages of the ϕ_{ST} estimates (with negative

estimates set to zero) are 0·151 for Surin, Trengganu and Bogor, 0·097 for the Philippines, Sabah, Sarawak and Sulawesi, but 0·220 between these two groups of populations. Thirdly, the pattern of differentiation (splitting) in the tree derived from d_{XY} distances (Fig. 5) shows a clear clustering of the Philippines and eastern islands populations, which is separate from the mainland populations (Surin and Trengganu). These are in turn more closely related to the river buffalo populations.

Overall, our results support the following hypothesis for the evolution of the water buffalo: (i) that the species originated in mainland south-east Asia, (ii) that the species then spread north to China, and west to the Indian subcontinent, where the river type evolved, (iii) following domestication in China, the domesticated swamp buffalo spread with rice cultivation (a) through Taiwan, to the Philippines and to the eastern islands of Borneo and Sulawesi, and (b) south through mainland south-east Asia (likely interbreeding with wild buffalo) to Peninsular Malaysia and on to the western islands of Indonesia (Sumatra and Java).

Acknowledgements

We gratefully acknowledge the co-operation and assistance during sample collection of Professor C Chantalakhana, Ms Ancharlee and the Thai Department of Livestock Development; Professor R L Anacleto, Central Mindanao University of the Philippines; Dr P Abeynayake, University of Peradeniya, Sri Lanka; Dr Sudarisman and Dr P Daniels, BALITVET, Indonesia; Dr S P Chia, Department of Animal Services, Sabah; Dr K Y Hsiung, Department of Agriculture, Sarawak; Dr M Hilmi, Universiti Putra Malaysia; Mr O S Selvaraj, University of Malaya; and Ms J Purdie, Department of Primary Industry and Fisheries, NT, Australia. This work was supported by the Australian Centre for International Agricultural Research, the Malaysian Program for Intensification of Research in Priority Areas (Project no. 1-07-05-057), and Universiti Putra Malaysia.

References

Amano T. (1983) Genetic differences between swamp and river buffaloes in biochemical and immunological characteristics. In: *Proceedings of the Preconference Symposium of the 5th World Conference on Animal Production*, Tsukuba, Japan (ed. by H. Shimizu), pp. 131–5.·

Amano T., Namikawa T. & Natasasmita S. (1982) Blood protein polymorphisms of water buffaloes and anoas in Indonesia. In: *The Origin and Phylogeny of Indonesian Native Livestock*, Kagoshima, Japan. Part III (Report by Grant-in-Aid for Overseas Scientific Survey, no. 57043041), pp. 57–65.

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 (Suppl. 1),** 29–36.

Anderson S., de Bruijn M.H.L., 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.

Barker J.S.F., Tan S.G., Selvaraj O.S. & Mukherjee T.K. (1997a) Genetic variation within and relationships among populations of Asian water buffalo (*Bubalus bubalis*). *Animal Genetics* **28,** 1–13.

Barker J.S.F., Moore S.S., Hetzel D.J.S., Evans D., Tan S.G. & Byrne K. (1997b) Genetic diversity of Asian water buffalo (*Bubalus bubalis*): microsatellite variation and a comparison with protein-coding loci. *Animal Genetics* **28,** 103–15.

Bellwood P. (1992) Southeast Asia before history. In: *The Cambridge History of Southeast Asia*, Vol. 1 (ed. by N. Tarling), pp. 55–136. Cambridge University Press, Cambridge.

Bulmer M., Wolfe K.H. & Sharp P.M. (1991) Synonymous nucleotide substitution rates in mammalian genes: implications for the molecular clock and the relationship of mammalian orders. *Proceedings of the National Academy of Sciences, USA* **88,** 5974–8.

Chen Y.C. & 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. (1974) *The Husbandry and Health of the Domestic Buffalo* (ed. by W.R. Cockrill). FAO, Rome.

Excoffier L. (1995) *AMOVA 1·55 (Analysis of Molecular Variance).* http://acasun1.unige.ch/LGB/Software/Windoze/amova.

Excoffier L., Smouse P.E. & Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131,** 479–91.

Felsenstein J. (1993) *PHYLIP (Phylogeny Inference Package) version 3·5c.* Distributed by the author. Department of Genetics, University of Washington, Seattle, WA.

Fischer H. & Ulbrich F. (1968) Chromosomes of the Murrah buffalo and its crossbreds with the Asiatic swamp buffalo (*Bubalus bubalis*). *Zeitschrift für Tierzüchtung und Züchtungsbiologie* **84,** 110–4.

Hochberg Y. (1988) A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75,** 800–2.

Kimura M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16,** 111–20.

Kumar S., Tamura K. & Nei M. (1993) *MEGA: Molecular Evolutionary Genetics Analysis, version 1·01.* The Pennsylvania State University, University Park, PA.

Lau C.H., Yusoff K., Tan S.G. & Yamada Y. (1995) A PCR cycle sequencing protocol for population mitochondrial cytochrome b analysis. *BioTechniques* **18,** 262–5.

Lessios H.A. (1992) Testing electrophoretic data for agreement with Hardy–Weinberg expectations. *Marine Biology* **112,** 517–23.

Li W.-H., Gouy M., Sharp P.M., O'hUigin C. & Yang Y.-W. (1990) Molecular phylogeny of rodentia, lagomorpha, primates, artiodactyla, and carnivora and molecular clocks. *Proceedings of the National Academy of Sciences, USA* **87,** 6703–7.

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, USA* **91,** 2757–61.

Macgregor R. (1939) *The domestic buffalo*. Thesis presented to the Royal College of Veterinary Surgeons (quoted by I.L. Mason in Cockrill (1974), 1–47).

Mason I.L. (1974) Species, types and breeds. In: *The Husbandry and Health of the Domestic Buffalo* (ed. by W.R. Cockrill), pp. 1–47. Food and Agriculture Organization of the United Nations, Rome.

McElroy D., Moran P., Bermingham E. & Kornfield I. (1991) *REAP The restriction enzyme analysis package. Version 4·0*. Department of Zoology, Migratory Fish Research Institute and Centre for Marine Studies, University of Maine, ME.

Nei M. (1985) Human evolution at the molecular level. In: *Population Genetics and Molecular Evolution* (ed. by T. Ohta & K. Aoki), pp. 41–64. Springer-Verlag, Berlin.

Nei M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M. & Tajima F. (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics* **97,** 145–63.

Prim R.C. (1957) Shortest connection networks and some generalizations. *Bell System Technical Journal* **36,** 1389–401.

Raymond M. & Rousset F. (1995a) An exact test for population differentiation. *Evolution* **49,** 1280–3.

Raymond M. & Rousset F. (1995b) GENEPOP (Version 1·2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86,** 248–9.

Raymond M. & Rousset F. (1997) *GENEPOP (Version 3·1)*. ftp, ftp.cefe.cnrs-mop.fr, directory pub/pc/msdos/genepop.

Rohlf F.J. (1988) *NTSYS-pc. Numerical Taxonomy and Multivariate Analysis System. Version 1·50*. Exeter Publications, New York.

Saitou N. & Nei M. (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4,** 406–25.

Stoneking M., Sherry S.T., Redd A.J. & Vigilant L. (1992) New approaches to dating suggest a recent age for the human mtDNA ancestor. *Philosophical Transactions of the Royal Society, London B* **337,** 167–75.

Takahata N. & Nei M. (1985) Gene genealogy and variance of interpopulational nucleotide differences. *Genetics* **110,** 325–44.

Tanaka K., Solis C.D., Masangkay J.S., Maeda K., Kawamoto 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.

Tanaka K., Yamagata T., Masangkay J.S. *et al.* (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.

Thompson J.D., Higgins D.G. & Gibson T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22,** 4673–80.

Ulbrich F. & Fischer H. (1967) The chromosomes of the Asiatic buffalo (*Bubalus bubalis*) and the African buffalo (*Cyncerus caffer*). *Zeitschrift für Tierzüchtung und Züchtungsbiologie* **83,** 219–23.