MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 1530–1550 doi: 10.1111/mec.13518

Strong and stable geographic differentiation of swamp buffalo maternal and paternal lineages indicates domestication in the China/Indochina border region

YI ZHANG,* YONGFANG LU,* MARNOCH YINDEE,† KUAN-YI LI,‡ HSIAO-YUN KUO,§ YU-TEN JU,‡ SHAOHUI YE,¶ MD OMAR FARUQUE,** QIANG LI,* YACHUN WANG,* VU CHI CUONG,†† LAN DOAN PHAM,†† BOUNTHONG BOUAHOM,‡‡ BINGZHUANG YANG,§§ XIANWEI LIANG,§§ ZHIHUA CAI,¶¶ DIANNE VANKAN,*** WALLAYA MANATCHAIWORAKUL,††† NONGLID KOWLIM,††† SOMPHOT DUANGCHANTRASIRI,‡‡‡ WORAWIDH WAJJWALKU,††† BEN COLENBRANDER,§§§ YUAN ZHANG,* PETER BEERLI,¶¶¶ JOHANNES A. LENSTRA§§§ and J. STUART F. BARKER****

*National Engineering Laboratory for Animal Breeding, Key Laboratory of Animal Genetics and Breeding and Reproduction of MOA, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China, †Department of Clinical Science and Public Health, Faculty of Veterinary Science, Mahidol University, Kanchanaburi Campus, Kanchanaburi 71150, Thailand, ‡Department of Animal Science and Technology, National Taiwan University, Taipei 10673, Taiwan, §Livestock Research Institute, Council of Agriculture, Tainan 71246, Taiwan, ¶College of Animal Science and Technology, Yunnan Agricultural University, Kunming 650201, China, **Department of Animal Breeding and Genetics, Bangladesh Agricultural University, Mymensingh 2202, Bangladesh, ††Key Laboratory of Animal Cell Technology, National Institute of Animal Sciences, Tu Liem, Hanoi 100000, Vietnam, ‡‡National Agriculture and Forestry Research Institute, P.O. Box 811, Vientiane Capital, Lao P.D.R., §§Guangxi Buffalo Research Institute, Chinese Academy of Agriculture Sciences, Nanning 530001, China, ¶¶College of Animal Science, Anhui Science and Technology University, Fengyang 233100, China, ***The School of Veterinary Science, University of Queensland, Gatton Campus, Gatton, Qld 4343, Australia, †††Department of Pathology, Faculty of Veterinary Medicine, Kasetsart University, Kamphaengsaen, Nakhon Pathom 73140, Thailand, ‡‡‡Khao-Nang-Ram Wildlife Research Station, Department of National Parks, Wildlife and Plant Conservation, Bangkok 10900, Thailand, §§§Faculty of Veterinary Medicine, Utrecht University, Yalelaan 104, 3584 CM, Utrecht, The Netherlands, ¶¶¶Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, USA, ****School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia

Abstract

The swamp type of the Asian water buffalo is assumed to have been domesticated by about 4000 years BP, following the introduction of rice cultivation. Previous localizations of the domestication site were based on mitochondrial DNA (mtDNA) variation within China, accounting only for the maternal lineage. We carried out a comprehensive sampling of China, Taiwan, Vietnam, Laos, Thailand, Nepal and Bangladesh and sequenced the mtDNA *Cytochrome b* gene and control region and the Y-chromosomal *ZFY*, *SRY* and *DBY* sequences. Swamp buffalo has a higher diversity of both maternal and paternal lineages than river buffalo, with also a remarkable contrast between a weak phylogeographic structure of river buffalo and a strong geographic differentiation of swamp buffalo. The highest diversity of the swamp buffalo maternal lineages was found in south China and north Indochina on both banks of the Mekong River, while the highest diversity in paternal lineages was in the China/Indochina border region. We propose that domestication in this region was later followed by introgressive capture of wild cows west of the Mekong. Migration to the north followed the Yangtze valley as well as a more eastern route, but also involved translocations of both

Correspondence: Yuan Zhang, Fax:+86-10-62733687;
E-mail: changy@cau.edu.cn

cows and bulls over large distances with a minor influence of river buffaloes in recent decades. Bayesian analyses of various migration models also supported domestication in the China/Indochina border region. Coalescence analysis yielded consistent estimates for the expansion of the major swamp buffalo haplogroups with a credibility interval of 900 to 3900 years BP. The spatial differentiation of mtDNA and Y-chromosomal haplotype distributions indicates a lack of gene flow between established populations that is unprecedented in livestock.

Keywords: *Bubalus bubalis*, migration models, mitochondrial DNA, Y-chromosomal sequences

Received 21 April 2015; revision received 30 November 2015; accepted 10 December 2015

Introduction

Domestications of crop plants and animals from 10 000 to 12 000 years BP (before present) were crucial events in the history of human civilizations (Diamond 1997; FAO 2007). In the past, archaeological findings have been invoked to suggest the place and time of domestication of various animal species, but molecular genetic methods (Lenstra *et al.* 2012) are now being used to provide information on the ancestral species, domestication site and subsequent demographic history of cattle (Loftus *et al.* 1994; Bradley *et al.* 1996; Beja-Pereira *et al.* 2006; Ajmone-Marsan *et al.* 2010; Chen *et al.* 2010; Bollongino *et al.* 2012; Decker *et al.* 2014), horses (Cieslak *et al.* 2010; Warmuth *et al.* 2012), sheep (Tapio *et al.* 2006; Meadows *et al.* 2007; Chessa *et al.* 2009; Kijas *et al.* 2012), goats (Cañón *et al.* 2006; Naderi *et al.* 2008; Nomura *et al.* 2013) and pigs (Giuffra *et al.* 2000; Fang & Andersson 2006). Groeneveld *et al.* (2010) have reviewed molecular studies of livestock domestication, while Wiener & Wilkinson (2011) surveyed the literature on gene mutations after domestication or accompanying breed development. For several species, it is becoming clear that domestication was not a single event and was followed by introgression of wild animals (Luikart *et al.* 2001; Bruford *et al.* 2003; Beja-Pereira *et al.* 2006; Chen *et al.* 2010; Larson & Fuller 2014).

The Asian water buffalo (*Bubalus bubalis*) is an important livestock animal as a source of food and draught power in tropical and subtropical regions, with more people in the world depending on it for their livelihoods than on any other domestic animal species (FAO 2000; http://faostat3.fao.org). Morphological and behavioural traits differentiate two types (Macgregor 1941) – the dairy river buffalo of the Indian subcontinent and west to the Balkans, Italy and Egypt, and the swamp buffalo, whose main use is as a draft animal in the region ranging from Assam (India) in the west through South-East Asia to the Yangtze valley of China in the east. The two types have been shown to be genetically distinct on the basis of chromosome number (river 2n = 50, swamp 2n = 48 – Ulbrich & Fischer 1967; Fischer & Ulbrich 1968), allozymes (Amano *et al.* 1980; Barker *et al.* 1997a) and autosomal as well as sex-linked DNA markers (Barker *et al.* 1997b; Zhang *et al.* 2006; Kumar *et al.* 2007a; Yindee *et al.* 2010). Archaeological, historical and anatomical evidence support a descent of both river and swamp domestic buffalo from the wild Asian water buffalo – *Bubalus arnee* (Cockrill 1984). The time of divergence of the two types has been estimated as 10 000–1 700 000 years BP, with 128 000–270 000 as the most likely range (Kumar *et al.* 2007a) and according to all estimates well before domestication (Zhang *et al.* 2011).

Genetic data clearly indicate independent domestications of river and swamp buffalo (Lau *et al.* 1998; Kumar *et al.* 2007a; Lei *et al.* 2007; Yindee *et al.* 2010). Domestication of the river type in the western region of the Indian subcontinent has been estimated to have occurred 6300 years BP (Kumar *et al.* 2007b), but the place and time of domestication of the swamp buffalo are controversial. Cockrill (1984) suggested that the swamp buffalo was domesticated in the Yangtze valley 4000–5000 years ago. However, Holocene remains of water buffalo in China have now been attributed to the wild *Bubalus mephistopheles* rather than *B. arnee* (Wang & Zhang 2011). Patel & Meadow (1998) suggested that *B. arnee* ranged 'perhaps as far east as the coast of southern China', and this is compatible with its known distribution in Indochina within historical times. In an earlier text, Epstein (1969) noted that domesticated buffaloes were present in China by the time of the Shang dynasty (ca. 1766–1123 BC), or even earlier, and suggested that they were introduced from South-East Asian areas bordering China. Recent mtDNA and microsatellite analyses (Yindee 2010; Zhang *et al.* 2011) and bones from domestic buffalo (2300–500 BC, Ban-Chiang, northern Thailand (Higham 2002) suggest swamp buffalo domestication in the region of south China, north Thailand and Indochina. In contrast, Yue *et al.* (2013) proposed southwest China as the domestication site, but their study included Chinese populations only.

In order to localize more accurately the domestication site of swamp buffalo and to reconstruct its subsequent dispersal over the present distribution range, we report

a large-scale sampling and analyses of mitochondrial DNA from 1100 water buffaloes (male and female) sampled from 43 populations in China, Taiwan, Vietnam, Laos, Thailand, Nepal and Bangladesh and an analysis of Y-chromosomal DNA from 495 male animals. Our analysis shows a dispersed distribution of the mtDNA haplogroups of the river buffalo, as commonly observed for domestic species, but unusually strong contrasts in the distributions of swamp buffalo haplotypes for both mtDNA and Y-chromosomal DNA. The haplogroup distributions and the patterns of gene flow suggest domestication in the China/Indochina border region, local recruitment of wild cows in Indochina west of the Mekong River and a much lower gene flow between local populations than that observed for other livestock species, including the closely related river buffalo.

Materials and methods

Sampling and DNA sequencing

We used samples of 1100 buffaloes (males and females) for mtDNA analysis and samples of 495 bulls, including 23 previously reported Thai animals (Yindee *et al.* 2010) for Y-chromosomal analysis. Locations of the populations and population codes are given in Table 1 and Fig. 1. Most populations from which animals were sampled were either swamp or river type. However, four of the populations sampled were known to have mixed river/swamp ancestry: BD_C is primarily a river buffalo population with swamp introgression; BD_E and BD_SE are primarily swamp buffaloes with river introgression; and C_BL was originally a swamp population, but is now primarily river buffalo due to extensive introgression since at least the early 1900s. These populations were classified as either swamp or river according to the predominant type. The sequences obtained from animals in these populations were clearly swamp or river, and the number of sequences (both mtDNA and Y chromosome) obtained for each population is given in Table 1. Most animals were sampled randomly from village populations, while minimizing family relationships on the basis of information supplied by the owner. Exceptions to this were the Nepal samples (see Flamand *et al.* 2003 and Zhang *et al.* 2011 for details), and the samples from the C_BH, C_HZ and TW populations, which are herds kept for the purpose of conservation. A few bulls of the Chinese swamp buffalo populations C_DS, C_FY and C_ER were found to carry river Y-chromosomal DNA (Table 1). In all analyses of both mitochondrial and Y-chromosomal DNA, river and swamp samples were considered separately.

DNA extraction and/or sequencing were performed in Beijing (samples from China, Vietnam, Laos, Nepal and Bangladesh), Kamphangsaen (samples from Thailand) and Taipei (samples from Taiwan). Genomic DNA was isolated from ear skin tissue or blood cells using a standard phenol–chloroform procedure, except for samples from Laos, where DNA was extracted from hair bulbs with Genomic DNA Magnetic-bead Extraction kit (TIANGEN, Beijing).

PCR (primers listed in Table S1, Supporting information) was carried out in Beijing and Taipei with 50 ng genomic DNA in 20 μL standard PCR buffer with 2 mM $MgCl_2$, 0.2 mM of each dNTP, 10 μM of each primer and 1 U of Taq DNA polymerase. After 10 min at 95 °C, 30 cycles were performed at 95 °C for 30 s, 55–63 °C for 30 s (Table S1, Supporting information) and 72 °C for 45 s, followed by a final extension step of 72 °C for 5 min. In Kamphaengsaen, PCR was carried out with 50 ng genomic DNA in 25 μL standard PCR buffer with 1.5 mM MgCl2, 0.2 mM of each dNTP, 10 μM of each primer and 1 U of Taq DNA polymerase. After 3 min at 95 °C, 35 cycles were performed at 95 °C for 15 s, 55–60 °C for 30 s (Table S1, Supporting information) and 72 °C for 45 s, followed by a final extension step of 72 °C for 5 min. Dideoxy sequencing was performed using the BIGDYE™ Terminator Cycle Sequencing Kit (Applied Biosystems, Foster, CA, USA). Raw sequence fragments were assembled and checked with the program SEQMAN (DNAstar, Madison, WI). Sequences were aligned using MAFFT online version 7 (http://mafft.cbrc.jp/alignment/server/). Insertion/deletions in aligned sequences (Fig. S1, Supporting information) were excluded from the analyses. Amplification of nuclear mtDNA insertions has been excluded by (1) amplification of both control region and *Cytochrome b* gene on five reference samples in Beijing with the primers used in Kamphangsaen, which yielded sequences identical to those obtained with the primers used in Beijing, and (2) comparing the phylogenies of *Cytochrome b* and control region fragments, respectively, which were entirely consistent.

The neighbouring mtDNA control region (889 bp) and the *Cytochrome b* gene (1140 bp) were analysed by dideoxy sequencing of PCR products. For Y-chromosomal analysis, 5505 bp of the Y-chromosomal *ZFY*, *SRY* and *DBY* genes from 79 randomly selected bulls were sequenced (Nijman *et al.* 2008). This revealed SNPs in the fragments *ZFY1*, *ZFY3*, *SRY5* and *DBY1* (altogether 2310 bp), which were sequenced in the remaining 386 male samples. For haplotype distribution and diversity analysis, 23 published Thailand sequences (Yindee *et al.* 2010) were added to the data set.

Statistical analyses

We have constructed phylogenetic trees and calculated several summary statistics, which describe the genetic

Table 1 Populations, codes, sample sizes and geographical coordinates

			Sample size				Geographical coordinates	
			Swamp		River			
Country/region	Population	Code	mtDNA	Y	mtDNA	Y	Latitude	Longitude
Swamp buffalo populations								
China	Dangshan*	C_DS	36	3		3	34.40	116.58
	Shannan*	C_SN	24				33.16	107.03
	Haizi	C_HZ	25	3			32.89	120.33
	Shanqu	C_SQ	25	5			33.00	118.50
	Fengyang	C_FY	53	6		2	32.86	117.57
	Dongliu	C_DL	26	19			30.10	117.05
	Jianghan	C_JH	25	1			30.36	112.19
	Binhu	C_BH	25	10			29.51	112.55
	Enshi*	C_ES	25	29			29.72	109.32
	Guizhou	C_GZ	25	41			27.95	107.72
	Yanjin	C_YJ	24	18			28.11	104.24
	Fuan	C_FA	25	21			27.10	119.65
	Xiajiang	C_XJ	24	1			27.58	115.32
	Xinfeng	C_XF	19	12			25.39	114.93
	Fuzhong	C_FZ	25	30			24.53	111.30
	Xinglong	C_XG	23	18			18.75	110.22
	Dechang*	C_DC	24	30			27.65	102.21
	Xilin	C_XL	22	21			24.49	105.09
	Diandongnan	C_DD	25	29			23.02	104.40
	Dali	C_DA	25	11			26.35	100.20
	Ershan	C_ER	25	26		1	24.33	102.20
	Dehong	C_DH	39	25			24.43	98.59
Taiwan	Taiwan	TW	29	7			23.95	121.55
Vietnam	Ha Giang (North)	VN_N	75	26			22.08	104.33
	Tay Ninh (South)	VN_S	25	7			11.20	106.10
Laos	Xayaboury (North)	LA_N	39	8			19.15	101.45
	Khammouane (Central)	LA_C	17	5			17.30	105.20
	Champasak (South)	LA_S	40	8			14.53	105.52
Thailand*†	North	TH_N	11	1			19.84	99.78
	Northeast	TH_NE	18	12			15.00	103.36
	Central	TH_C	20	5			13.75	101.15
	South	TH_S	5	3			9.91	99.11
	Huai Kha Khaeng sanctuary, Uthai Thani	TH_Wild		2			15.42	99.23
Bangladesh	Bangladesh East	BD_E	16	4	5	8	24.90	91.85
	Bangladesh Southeast	BD_SE	2		2		21.71	92.40
	Total		886	447	7	14		
Swamp buffaloes for mtDNA analysis: 893								
Swamp buffaloes for Y-chromosomal analysis: 461								
River buffalo populations								
China	Binlangjiang	C_BL	24		26	23	25.45	98.33
Bangladesh	Bangladesh Central	BD_C	3	3	22	1	24.59	90.40
	Bangladesh West	BD_W			48	4	24.37	88.60
Nepal	Nepal Wild	NP_W			7	3	26.74	86.98
	Nepal Hybrid	NP_H			15		26.74	86.98
	Nepal Domestic	NP_D			20		26.74	86.98
China	China Nili Ravi	C_NR			20		22.90	108.35
	China Murrah	C_MR			22		22.90	108.35
	Total		27	3	180	31		

Table 1 *Continued*

Country/region	Population	Code	Sample size				Geographical coordinates	
			Swamp		River			
			mtDNA	Y	mtDNA	Y	Latitude	Longitude
River buffaloes for mtDNA analysis: 207 River buffaloes for Y-chromosomal analysis: 34 All buffalo populations All mtDNA samples: 1100 All Y samples: 495			913	450	187	45		

*Coordinates for C_DS, C_SN, C_ES, C_DC and the Thailand populations are averages over subpopulation localities – see Supporting Information, Supplement to Table 1.
†Y-chromosomal sequences of Thailand buffalo were from Yindee *et al.* (2010).

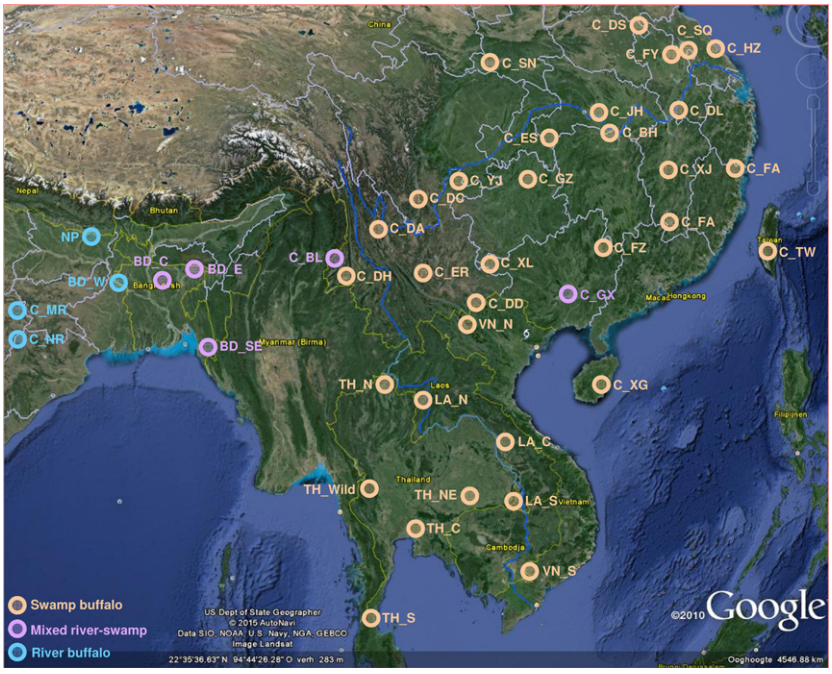

Fig. 1 Locations of sampled populations. For abbreviations, see Table 1. Population C_GX from Guangxi (Yue *et al.* 2013) is included to show the presence of hybrid river–swamp buffaloes in southern China. The imported Chinese river buffalo populations C_MR (from India) and C_NR (from Pakistan) are plotted in India.

variation in maternal and paternal genetic lineages of different regions and haplogroups as support for inferring historic gene flows. Except in the few cases noted, all analyses have been performed on the combined *Cytochrome b* and control region sequences treated as one locus.

Diversity parameters such as haplotype diversity (H), nucleotide diversity with the Jukes–Cantor correction (π) and the average number of pairwise differences were calculated for each population, region or group (see definitions later, Fig. 2) using the program DNASP v5 (Librado & Rozas 2009), which was also used for performing neutrality tests (Fu & Li 1993). Allelic richness was calculated by rarefaction with a sample size of 11 (mtDNA) or 5 (Y-chromosomal DNA) using the program ADZE (https://web.stanford.edu/group/rosenberglab/adze.html). To evaluate the diversity statistics in a data set with several diverged haplogroups, we analysed 10 random selections of 20 animals from the diverse Thai populations. This gave standard deviations for π, H and allelic richness of 38%, 13% and 23%, respectively, of their range (the difference between their highest and the lowest values), indicating that in our data set, haplotype diversity (H) is the most robust diversity measure.

Exact tests of population differentiation were based on haplotype frequencies and pairwise F_{ST} values (Tamura & Nei 1993 molecular distance), with 1000 permutations for significance testing. Differentiation of populations, regions or groups was analysed by the sum of squares decomposition underlying the analysis of molecular variance (AMOVA). Differentiation of

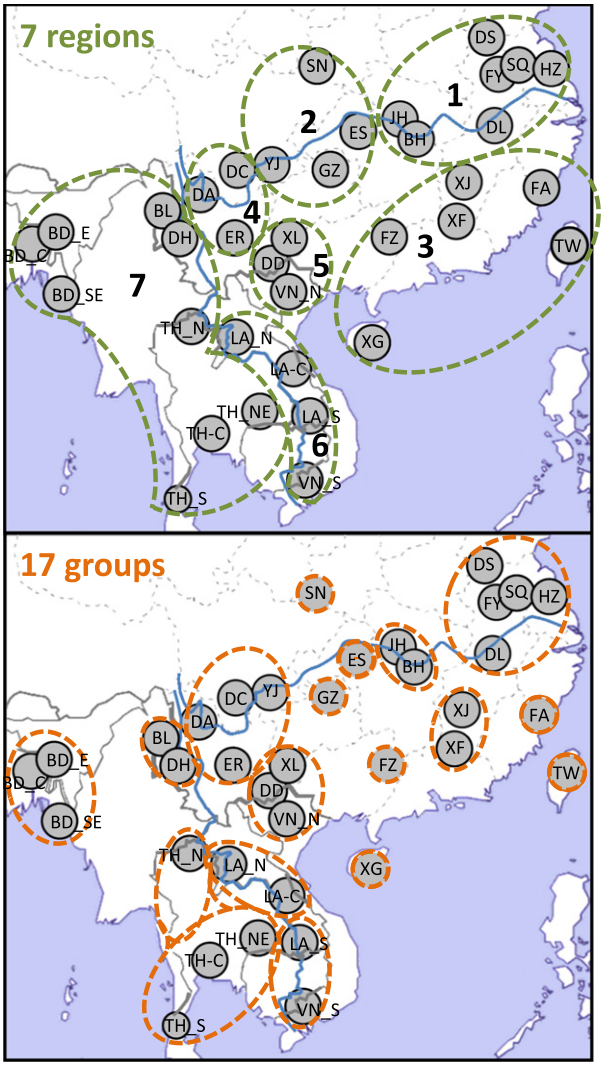

Fig. 2 Definition of 7 regions and 17 groups. For all Chinese populations, the prefix C_ has been omitted for clarity.

haplotypes was analysed by comparing mismatch distributions. Departure from equilibrium and neutrality was tested by calculations of the D (Tajima 1989) and Fs (Fu 1997) differentiation statistics, using 10 000 simulated samples for testing significance. These differentiation tests were carried out using ARLEQUIN v3.5 (http://cmpg.unibe.ch/software/arlequin3, Excoffier *et al.* 2005).

We constructed phylogenetic trees of mtDNA haplotypes to define the haplogroups. According to the Bayesian information criterion (BIC) implemented in the program MEGA5 (Tamura *et al.* 2011), the Kimura 2-parameter model with a gamma shape parameter (α) of 0.49 and fraction of evolutionary invariable sites of 0.76 was the most appropriate model for the combined swamp and river data. This model and parameters were then used in the MEGA5 program for construction of a maximum-likelihood tree. To assign individual sequences to haplogroups, neighbour-joining trees were constructed on the basis of the uncorrected number of changes between two sequences (*P* distance) using the program SPLITSTREE (Huson 1998). Geographic distributions of haplogroups were visualized by plotting pie plots on a topographical map and by median-joining networks drawn using the program NETWORK 4.6.1.1 (http://www.fluxus-engineering.com) following the recommendations of Bandelt *et al.* (1999).

Isolation by distance was tested for the swamp populations by calculation of pairwise F_{ST} estimated by ARLEQUIN (Tamura & Nei 1993 molecular distance) with negative values changed to zero and regression of pairwise $F_{ST}/(1 - F_{ST})$ values on the ln(geographic distance) between each pair of populations (Rousset 1997), using the ISOLDE program in the GENEPOP 3.4 package (Raymond & Rousset 1995), with significance of the correlation determined by a permutation test (Mantel 1967) with 10 000 permutations in both rows and columns.

For AMOVA (Table S2, Supporting information) and geographic trends in mtDNA and Y-chromosomal diversity (Table S9, Supporting information), the swamp populations were combined into 7 regions or 17 groups (Fig. 2), based on geographical proximity, regional topography and genetic differentiation. The seven regions were also used for surveying haplogroup diversity (Table S5, Supporting information), median-joining networks (Figs 4 and 7) and modelling of gene flow (Figs 9 and 10, Table S11, Supporting information). Population C_YJ has been assigned to the same group as the nearby populations of Region 4 because of a similar haplogroups distribution, but was placed in Region 2 (Fig. 2) because this maximized the proportion of the diversity assigned to the among-regions component in the AMOVA (Table S2, Supporting information).

Modelling of gene flow

We used three methods to analyse the spatial distribution of haplotypes:

1 The program SAMOVA clusters proximate populations while maximizing genetic differentiation (Dupanloup *et al.* 2002). It was not informative with our data set, as at most *k*-values (number of clusters) it generated clusters of a single population and one cluster of the remaining populations.
2 BAPS (Corander *et al.* 2008) optionally incorporates spatial information in a Bayesian model-based clustering algorithm on the basis of allele frequencies. However, with our data it reproduced only the division of the haplotypes into the haplogroups.

3 MIGRATE-N ver. 3.6.1 (Beerli 2006; Beerli & Palczewski 2010) investigates gene-flow models. It estimates the posterior distributions of (1) the mutation-scaled population size parameter θ, which is equivalent to $4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per site and generation; (2) the mutation-scaled migration rate M, which is m/μ, where m is the immigration rate per generation (Beerli 2006); and (3) marginal likelihoods of specified gene-flow models (Beerli & Palczewski 2010), but it does not estimate mutation rate parameters.

We applied MIGRATE-N to the data set of 913 combined swamp buffalo *Cytochrome b* and control region sequences. PAUP ver. 4.0b.10 (Swofford 2003) was used to estimate the transition–transversion ratio and the shape parameter of the gamma distribution for the mutation model F84 + G for input to MIGRATE-N – separately for the full data set, for the 716 haplogroup A (A1 and A2) sequences and the 572 haplogroup A1 sequences. Preliminary runs using the full migration model for the seven regions (Fig. 2; seven θ and 42 M parameter estimates) indicated the following run conditions: Bayesian inference; slice sampling; uniform priors for θ between 0 and 0.2 and for M between 0 and 25 000; static heating with temperatures 1.00, 1.50, 3.00 and 10^6; swapping among chains potentially occurring at every step; and 10 replicates. For each replicate, a burn-in of 10^4 steps was followed by 10^7, 3×10^7 or 5×10^7 parameter samplings recorded at intervals of 10^3. Criteria for convergence were an expected sample size (ESS) for all parameters $>10^3$ and a good agreement of mean and median estimates of all parameters. These criteria were met for all final reported runs. Convergence also was checked visually by plotting the recorded steps of the posterior values for all parameters. Marginal likelihoods of the data for each model were approximated using Bézier corrected thermodynamic integration (Beerli & Palczewski 2010). These marginal likelihoods were then used to calculate model probabilities (Burnham & Anderson 2002).

Demographic history

Skyline plots for each haplogroup were calculated using the program BEAST 2.3 (Bouckaert *et al.* 2014), using the Hasegawa–Kishino–Yano model with gamma site rate variation (three classes) while integrating over all possible shape parameter alpha values and all possible transition–transversion ratios. Site frequencies were calculated from the data. We used the random local clock model that allows different evolutionary rates among branches. Each haplogroup was run for 10 million steps. An effective sample size >100 was used as a sign of convergence. For details, see the example xml file deposited at Dryad doi: http://dx.doi.org/10.5061/dryad.16mp4.

Results

mtDNA sequences

Complete mitochondrial control region and *Cytochrome b* sequences were determined for 1100 Asian water buffaloes (GenBank Accession nos. KR007969–KR010168, Figs S1A, B, Supporting information, Table 1). We found in the control region two hypervariable mononucleotide tracts (Fig. S1C, Supporting information), which were disregarded in all subsequent analyses. Summary statistics of sequence variation for the swamp and river types are shown in Tables 2 and 3. Nucleotide diversity parameters of both the control region and the *Cytochrome b* sequences (Table 2, last three columns) indicate, in agreement with the ML and NJ trees (Figs 3, S2 and S3, Supporting information), that the swamp buffalo haplotypes are more variable than those of river buffalo. Both swamp and river mtDNA were found in the BD and C_BL populations of mixed swamp/river origin (Table 1). This has been found previously for a southeast Chinese population in Guangxi (Yue *et al.* 2013).

Analysis of the 913 swamp buffalo mtDNA sequences showed 198 SNPs defining 263 haplotypes (Table 3). The separation of the haplogroups was supported by bootstrap percentages of >50%, except for SA1 and SA2. This reflects the presence of a few SA1-SA2 intermediate sequences, but SA1 and SA2 are clearly differentiated in the median-joining network (Fig. 4). Phylogenetic analysis revealed five main and three rare haplogroups (Table 3, Figs 3, S2 and S3, Supporting information). The predominant haplotypes (SA1-001 and SA2-001) occurred 259 and 50 times, respectively (Table 3). The 187 river sequences had 68 SNPs, three haplogroups and 61 haplotypes, one of which (R1-01) occurred 30 times (Tables 2 and 3).

Diversity values (Table 2) and mismatch distributions (Fig. S4, Supporting information) indicated that the major swamp haplogroup A1 expanded later than A2, B1 and B2. The relatively rare haplogroup B2 is the oldest, while B3 seems to have emerged relatively recently. Regional haplogroup distributions (Table S3, Supporting information) and the median-joining network (Fig. 4) show extensive geographic differentiation with (1) a lower frequency of A2 in northern populations; (2) an absence of B1 near the upper reaches of the Yangtze; (3) B2 found mainly in south China and west of the Mekong and (4) B3 present only in China and Vietnam. With only two exceptions, the rare haplogroups C, D and E are confined to Thailand

Table 2 Summary statistics of sampling, sequences and diversity

Locus	Swamp/ river	Length (bp)	%G+C	Ts/Tv* bias	Optimal substitution model	Gamma	*N* (samples)†	*N* (pop)‡	*N* (Ht)§	*N* (SNPs)¶	*N* (occurrences of haplotype)			*H***	π††	Mean *N* (pairwise differences)
											Most common	2nd most common	Singletons			
mtDNA																
Cytochrome b	Swamp	1140	44.03	19.37	HKY		913	36	46	59	525	136	24	0.637	0.00338	3.85
	River	1140	43.61	9.01	HKY		187	10	11	10	131	34	6	0.476	0.00075	0.858
Control region	Swamp	889	40.15	134.95	HKY+G+I	0.552	913	36	232	139	288	52	148	0.889	0.01469	13.06
	River	889	38.94	92.19	T92+G	0.050	187	10	56	58	32	19	30	0.940	0.00851	7.57
2 loci	Swamp	2029	42.32	76.45	HKY+G+I	0.075	913	36	263	198	259	50	170	0.909	0.00833	16.91
	River	2029	41.56	50.19	HKY+G	0.050	187	10	61	68	30	18	35	0.946	0.00415	8.43
Y chromosome																
ZFY-SRY-DBY	Swamp	5505					71	17	6	5	33	14	1	0.708	0.00023	1.264
	River	5504					8	2	1	0	8	0	0	0	0	0
ZFY1-ZFY3-	Swamp	2310					450	34	11	9	219	109	2			
SRY5-DBY1‡‡	River	2310					45	8	2	1	40	5	0			

*Ts, transition; Tv, transversion.
†Number of samples.
‡Number of populations (see Table 1).
§Number of haplotypes.
¶Number of SNPs.
**Haplotype diversity.
††Nucleotide diversity.
‡‡*ZFY1*, *ZFY3*, *SRY5*, *DBY1*: see Nijman *et al.* (2008).

Table 3 Swamp and river buffalo mtDNA haplogroups

Haplogroup	*N* (samples)	*N* (Ht)	Predominant haplotypes		*N* (SNPs)	*H*	Mean *N* (diff)*	π
			Haplotype	Occurrences				
SA1	572	151	SA1-001	259	110	0.788	1.448	0.00071
SA2	144	50	SA2-001	50	65	0.865	2.336	0.00115
SB1	89	31	SB1-01	37	28	0.818	2.313	0.00114
SB2	37	10	SB2-02	14	13	0.820	2.694	0.00133
SB3	62	16	SB3-01	44	17	0.498	0.673	0.00033
SC	5	2			2	0.400	0.800	0.00039
SD	3	2			1	0.667	0.667	0.00033
SE	1	1						
Subtotal	913	263						
R1	139	46	R1-01	30	47	0.921	3.164	0.00156
R2	35	10	R2-01	18	10	0.718	1.082	0.00053
R3	13	5	R3-05	5	10	0.808	4.436	0.00219
Subtotal	187	61						

*Mean number of pairwise differences.
Other abbreviations are as in Table 2.

(Table S3, Supporting information). These findings are supported by haplogroup distributions of populations (Fig. 5), which are congruent with previous data (Yue *et al.* 2013, Fig. S5, Supporting information) and which show similar patterns for most nearby populations.

Calculation of the Tajima D, Fu's Fs and Fu & Li D* and F* parameters for the different haplogroups yielded negative values (Table S4, Supporting information), which is in agreement with Yue *et al.* (2013). These values became statistically significant with more than about 50 segregating sites and are consistent with population expansions.

Statistical support for phylogeographic structure may be derived from three observations:

1 Differences among the seven regions (Table S5, Supporting information) were found to be significant;
2 AMOVAs (Table S2, Supporting information) – for all populations, 2.36% of the total variation was assigned to the among populations component; for 17 groups, 2.49% of the total variation was assigned to the among groups component; and for seven regions, 2.15% of the total variation was assigned to the among regions component; and
3 the regression of $F_{ST}/(1 - F_{ST})$ on ln(geographic distance) was significant (slope = 0.029 ± 0.006, $P = 9 \times 10^{-6}$).
4 Modelling of gene flow generates low likelihoods for panmixia.

Figure 6 shows a geographic plot of haplogroup diversity (*H*) of the swamp buffalo populations, calculated for both mtDNA loci and for *Cytochrome b* only (Table S6, Supporting information). The highest diversity was found in the populations on both sides of the China/Indochina border and west of the Mekong River in Thailand. In the northern populations, haplotype diversity based on the combined two loci was high relative to the values observed for *Cytochrome b*, which probably reflects variable mutation rate in the control region. This would also explain the relatively high *H* values for the control region of river buffalo (Table 2) and a low correlation ($r^2 = 0.505$) of the population *H* values calculated for both halves of the control region (data not shown).

Y-chromosomal sequences

Comparison of 2310-bp Y sequence (GenBank Accession nos. GQ259327–GQ259332, KT186376–KT186427) revealed 9 SNPs in 450 swamp buffalo bulls and one SNP in 45 river buffalo bulls. The swamp SNPs defined four major and seven rare haplotypes (Tables S7 and S8, Supporting information), which include five reported in a previous study (Yindee *et al.* 2010). The minor haplotypes YS2B, YS2C and YS3B plus YS3C were found only in the C_FZ, C_DC and C_GZ populations, respectively. Only one major (YR1) and one minor (YR2) haplotype were found in the river buffalo bulls (Table S8, Supporting information). In populations of mixed swamp/river origin, both river and swamp (BD_C, BD_E) or only river (C_BL) haplotypes were found, while the major river haplotype has been introgressed into the swamp populations C_DS, C_FY and C_ER).

A median-joining network (Fig. 7) and a geographic plot of Y-chromosomal haplotypes (Fig. 8) showed a strong geographic differentiation with (1) a lower frequency of YS1 in northern populations except for the

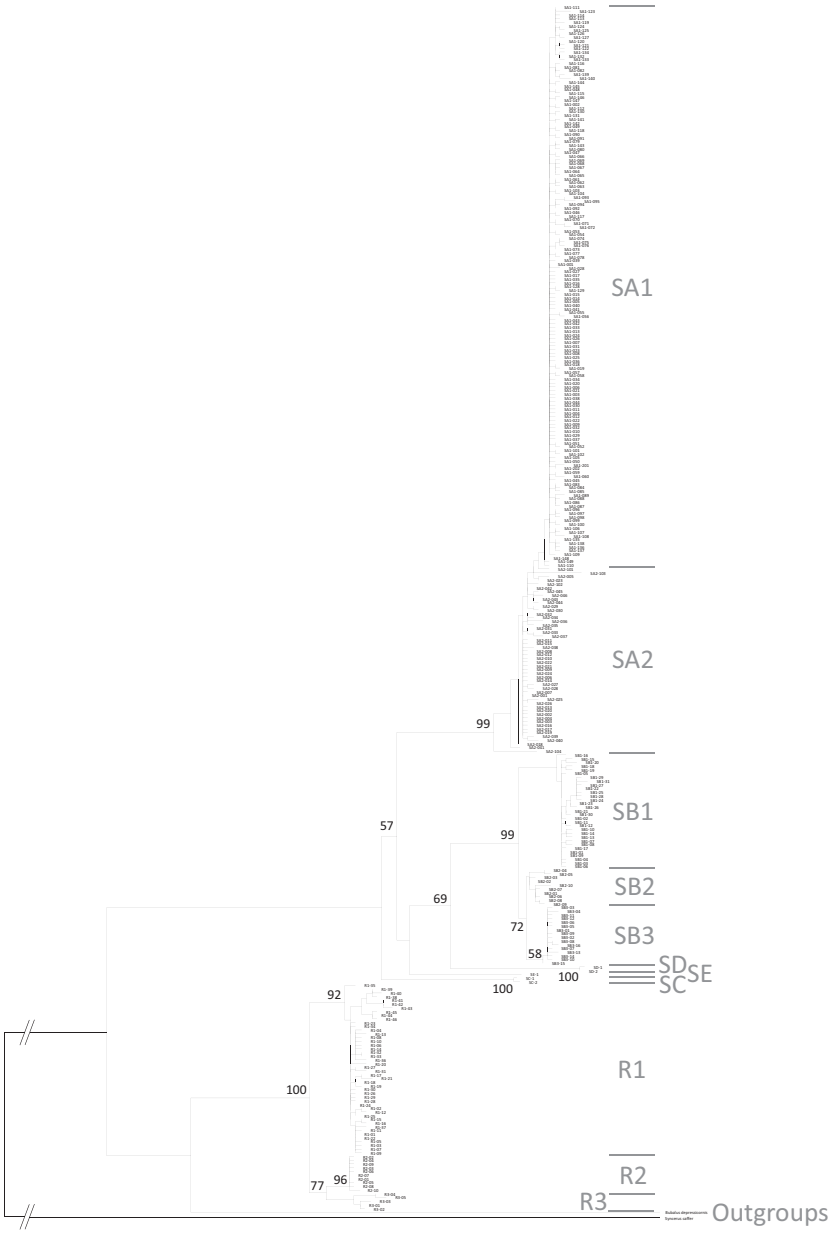

Fig. 3 Maximum-likelihood tree of water buffalo mtDNA haplotypes. Haplotype codes can be read by zooming in on the pdf file of Fig. S2 (Supporting information). SA1, SA2, SB1, SB2, SB3, SC, SD and SE indicate swamp buffalo haplogroups and R1, R2 and R3 river buffalo haplogroups. Anoa and African buffalo were used as out-groups. Bootstrap percentages larger than 50% for the haplogroups are shown.

lower Yangtze River region; (2) an absence of all YS2 haplotypes in southern Indochina; and (3) YS4A found mainly in Indochina and southwest China. The strong phylogeographic structure is supported by AMOVA (Table S2, Supporting information) with an appreciable proportion of the variation (5.56%) assigned to the among seven regions component. The highest diversity values (Tables S6 and S9, Supporting information) were generally found in populations east of the Mekong River.

Modelling of gene flow

Because of computational demands of the simulations, we selected, on the basis of qualitative interpretations of the haplogroup distributions, data sets that were simplified in different ways. First, we collapsed the populations to three main areas:

1 Area (1 2 3): regions 1, 2 and 3 in China, with haplogroups A1, A2, B1, B2 and B3 (Figs 2 and 9, Table S3, Supporting information);
2 Area (4): the mountainous Region 4 in southwest China with haplogroups A1, A2 and B3; and
3 Area (5 6 7): regions 5, 6 and 7 in south China, Indochina and Bangladesh with all observed haplogroups.

Seven migration models (Fig. 9) were evaluated in addition to the full model. Calculations carried out for

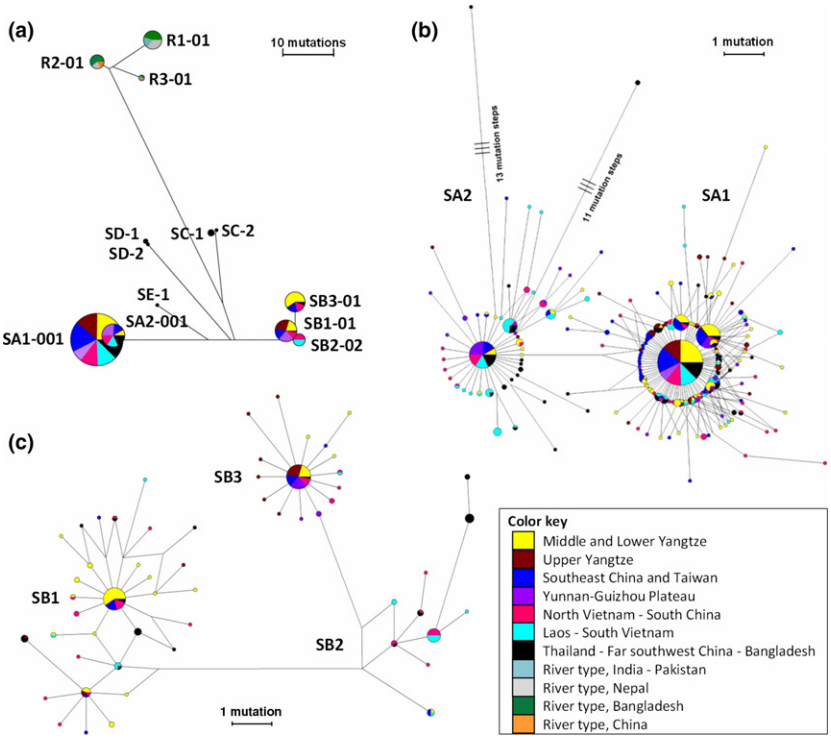

Fig. 4 Median-joining network of mtDNA haplotypes. (a) Most frequent haplotypes in river and swamp buffalo haplogroups; (b) swamp buffalo haplogroups SA1 and SA2; (c) swamp buffalo haplogroups SB1, SB2 and SB3. Circles indicate haplotypes; the area of the circles is proportional to the frequency of the corresponding haplotype and the coloured segments indicate the fraction of samples originating from the region specified by the colour key.

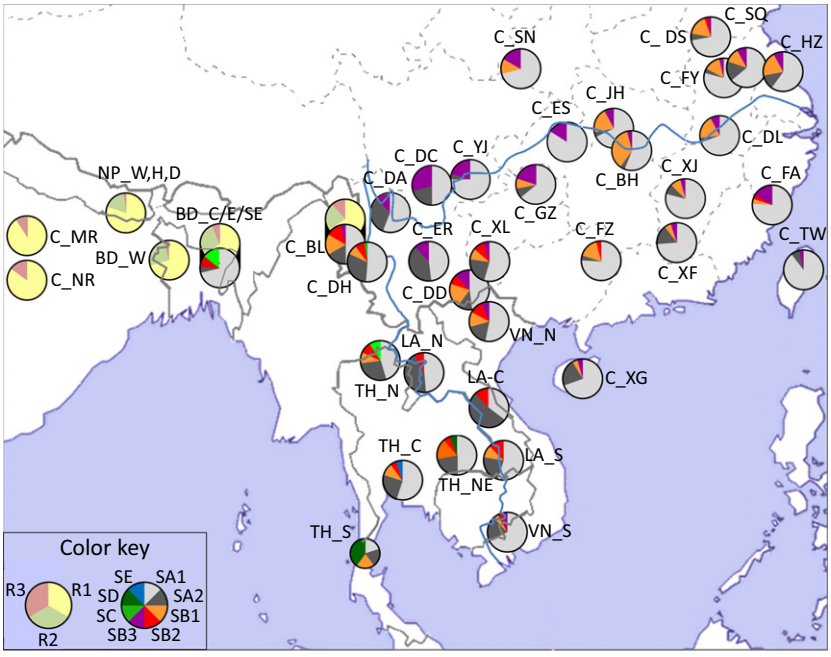

Fig. 5 mtDNA haplogroup distributions of river and swamp buffalo (Table S3, Supporting information). In the colour keys, SA1 to SE indicate swamp haplogroups, and R1 to R3 indicate river haplotypes.

all sequences, the 716 haplogroup A1 + A2 and the 572 haplogroup A1 sequences assigned the highest likelihood to models with Area (5 + 6 + 7) as the origin, but with different migration paths to Area (1 + 2 + 3) and Area 4 (Table S10, Supporting information).

Second, a more detailed analysis considered the seven regions defined in Fig. 2. In addition to the full migration model and a panmictic model, we chose 15 reduced models (Fig. 10). As suggested by the mtDNA and Y-chromosomal haplogroup distributions (Figs 5 and 8), these 15 models focus on Region 5 as the region of origin. Models 1A, 1B and 1C assume shortest distance migration radiating from Region 5, with mixing of populations along the Yangtze River (model 1B) or in

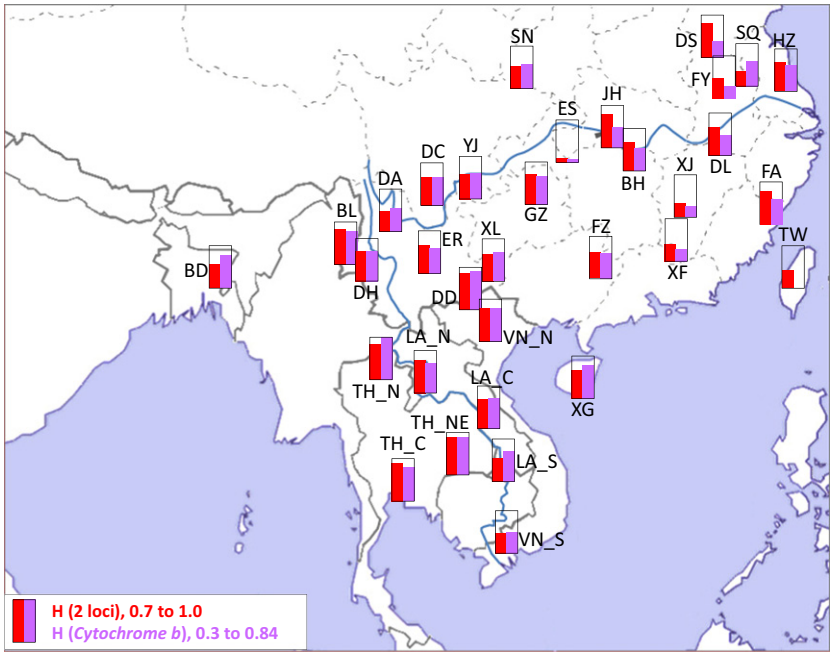

Fig. 6 mtDNA haplotype diversity of swamp buffalo populations. *H* values (Table S6, Supporting information) have been calculated for both *Cytochrome b* plus control region and for *Cytochrome b* only. The ranges of *H* values represented by the bars are indicated in the lower left corner. For all Chinese populations, the prefix C_ has been omitted for clarity. BD indicates the combined Bangladesh populations.

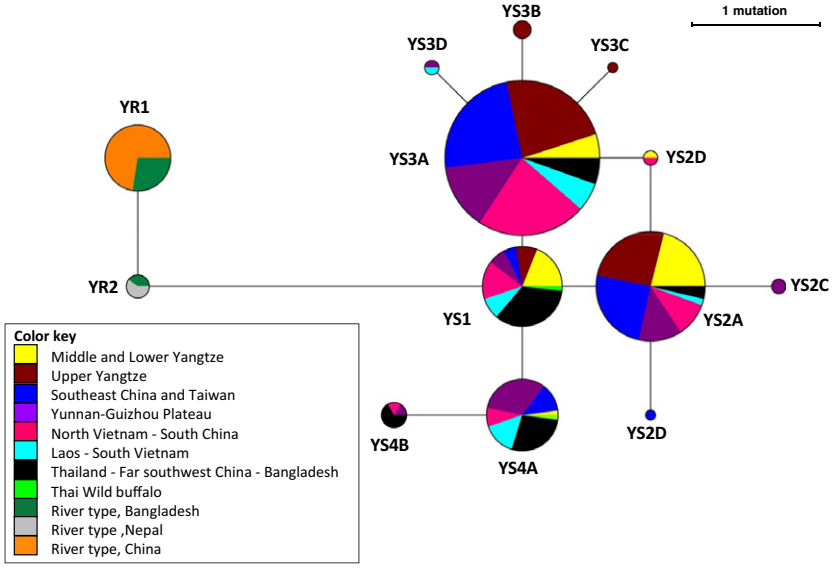

Fig. 7 Median-joining network of 13 Y-chromosomal haplotypes. Circles indicate haplotypes; the area of the circles is proportional to the frequency of the corresponding haplotype and the coloured segments indicate the fraction of samples originating from the region specified by the colour key.

the southeast lowlands (model 1C). Model 2 adds direct contact of regions 5 and 2. All other models except models 5 and 8 recognize that migration into the mountainous Region 4 would be easiest along the Yangtze River from Region 2. Model 3A is the simplest model with Region 5 as origin and Region 4 populated from Region 2. Model 3B also incorporates backflow from Region 1 to Region 3, while model 3C adds flow from Region 3 to Region 2. Model 4A allows migration from Region 6 to Region 7, while model 4B adds backflow from Region 1 to Region 3. Model 5 is a variant of 4A with direct contact of regions 4 and 5. Models 6A to 7B consider regions 6 and 7, respectively, as the original domestication site instead of or together with Region 5. Model 8 allows population of Region 4 across the Mekong from Region 7.

We used 3 random subsets of 40 individuals per region to select the seven most probable models (Table S11, Supporting information). In all three runs, the full and panmictic models and models 1B, 6A and 7A had the lowest likelihoods. Although the runs did not give a reproducible ranking of the other models, correlation coefficients of the marginal log likelihoods of -0.14, -0.26 and 0.56 allowed a consensus selection of the seven most probable models. Similar calculations with haplogroup A sequences generated high rankings

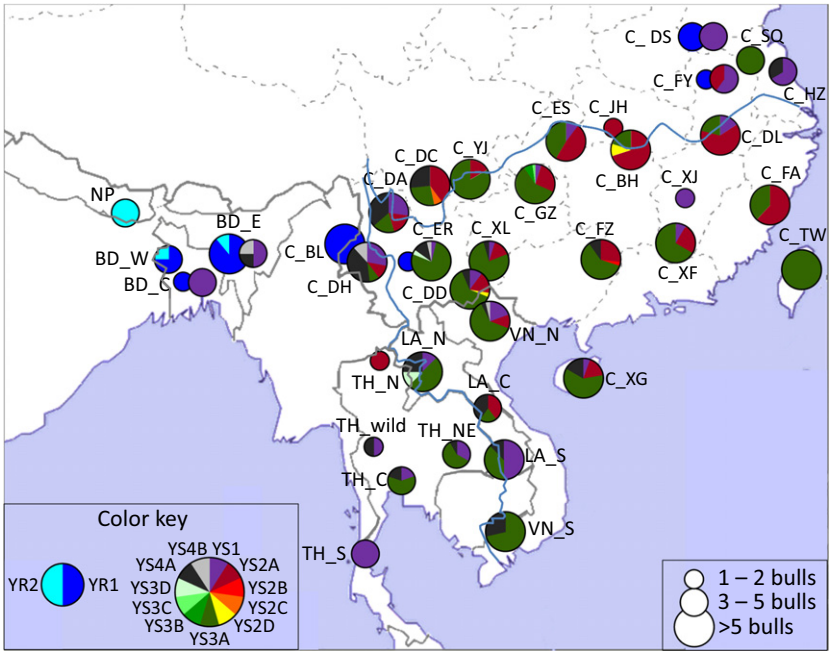

Fig. 8 Y-chromosomal haplogroup distributions of river and swamp buffalo populations (Table S8, Supporting information). The area of the circles is proportional to the number of bulls sampled from each population.

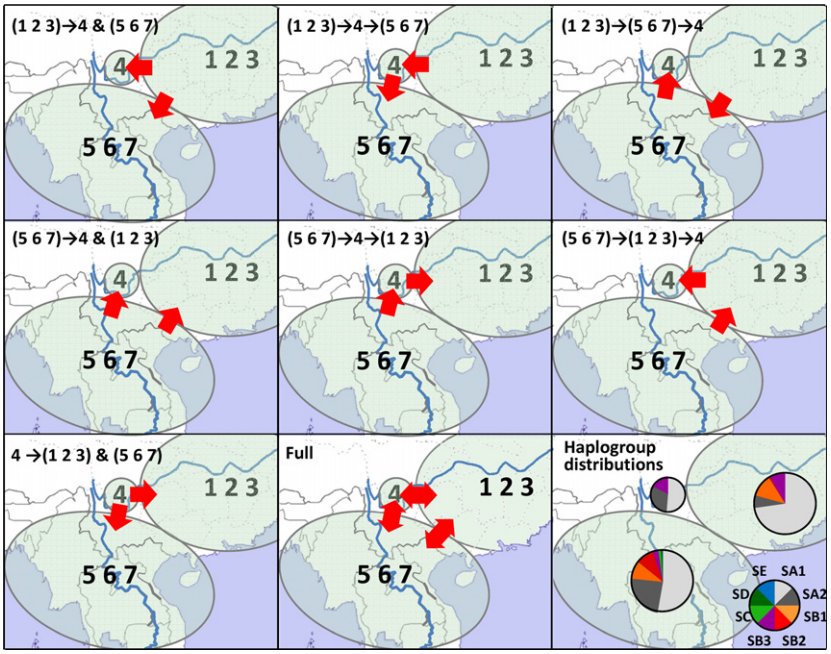

Fig. 9 Models of gene flow examined using the program MIGRATE-N, accounting for three main areas (see Fig. 2 for Regions 1 – 7). MtDNA haplogroup distributions are indicated.

for models 1A, 1B and 1C. These imply west-to-east gene flow along the Yangtze River, which is not compatible with the haplogroup distributions. The haplogroup A1 sequences also assigned relatively high ranking to models 1A and 1B, indicating a low power of these data sets with most mutations in the hypervariable control region. Thus, the haplogroup A and A1 data sets were not considered further.

Third, the seven models that ranked highest were rerun using all sequences and 5×10^7 parameter samplings (Table S12, Supporting information). This generated a ranking with model 4B as the most likely model. Model 2 implying an unlikely west-to-east gene flow along the Yangtze had the lowest ranking.

Finally, to explore migration models within a smaller region, possible pathways of gene flow among the 4 populations (C_XL, C_DD, VN_N from Region 5 and LA_N from Region 6) were tested for 11 models (Fig. 11). Model SC1 showed the best fit to the data (Table S13, Supporting information). Together with

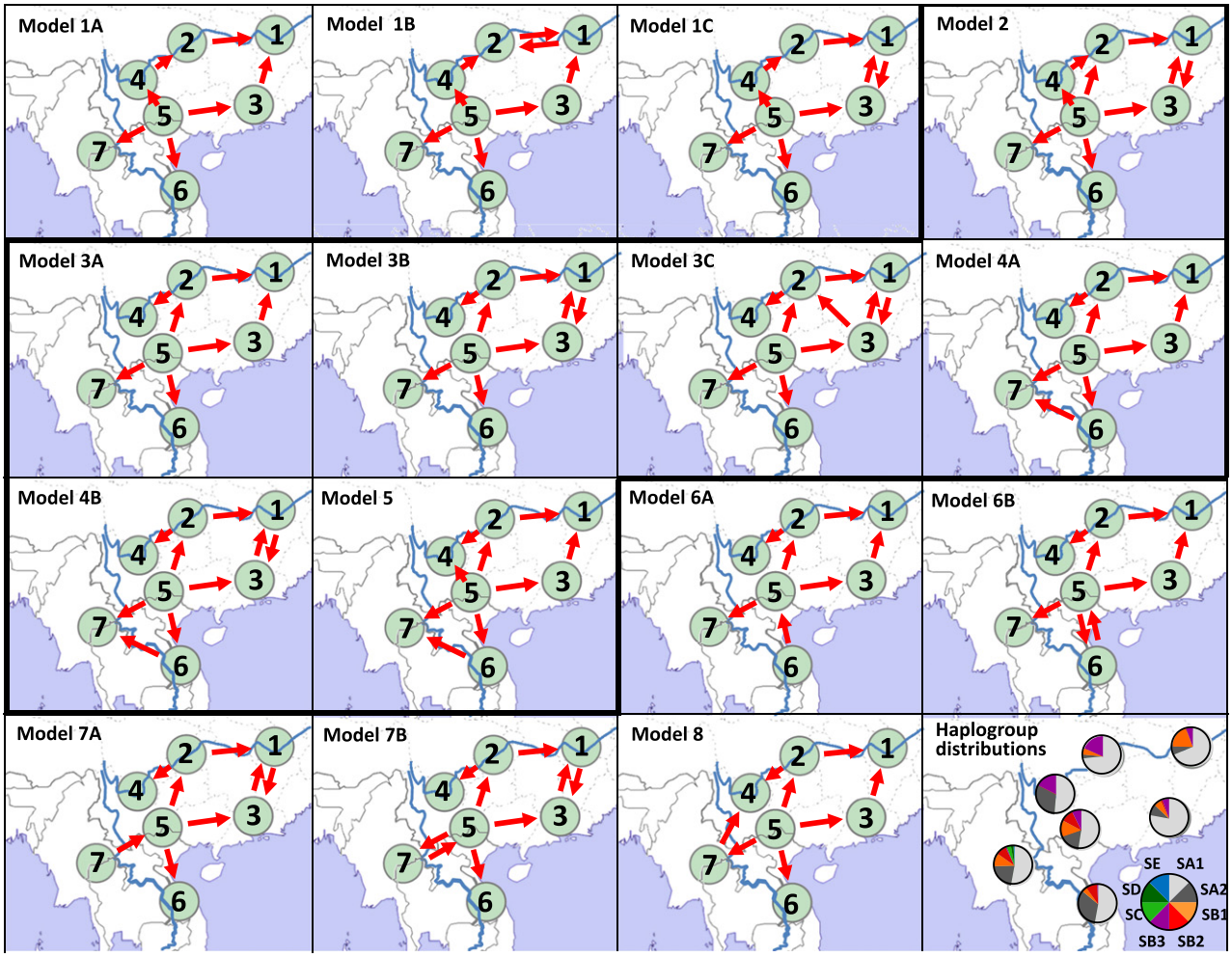

Fig. 10 Models of gene flow examined using the program MIGRATE-N, accounting for seven regions (Fig. 2). A bold line demarcates the seven models with the highest likelihood. MtDNA haplogroup distributions are indicated.

Region 5 as the most likely origin, this result suggests that the south China/Indochina border region area was the original domestication site.

Demographic history

The skyline plot (Fig. 12) of individual haplogroups shows a remarkably consistent increase in the population size for haplogroups SA1, SA2, SB1 and R1, a more recent expansion for haplogroups SB3 and R2 and no detectable or a much later expansion for haplogroups SB2 and R3. Assuming that domestication led to an expansion of the domestic population, we may derive estimates of the time of domestication from the first doubling in population size and converting the timescale in Fig. 12 to a time in years using mutation rates estimated for the Beringian steppe bison (Shapiro *et al.* 2004; 95% confidence interval 0.25 to 0.41/site/million year) corrected for using a larger mtDNA segment (see Table S14, Supporting information). Combining the upper and lower limits of both the 95% confidence intervals of the skyline plots (reflecting the uncertainty due to using a single locus) and the mutation rates leads for the major swamp buffalo haplogroup SA1 to an estimated time since the first expansion of 1107 to 3802 years BP (Table S14, Supporting information).

Discussion

On the basis of a comprehensive sampling and phylogenetic analysis, we have defined mtDNA and Y-chromosomal haplogroups. Haplogroup distributions and patterns of variation by region and haplogroup allow (i) to identify differences between the phylogeography of river and swamp buffalo, and (ii) to reconstruct and, for mtDNA, to model gene flows during and after domestication.

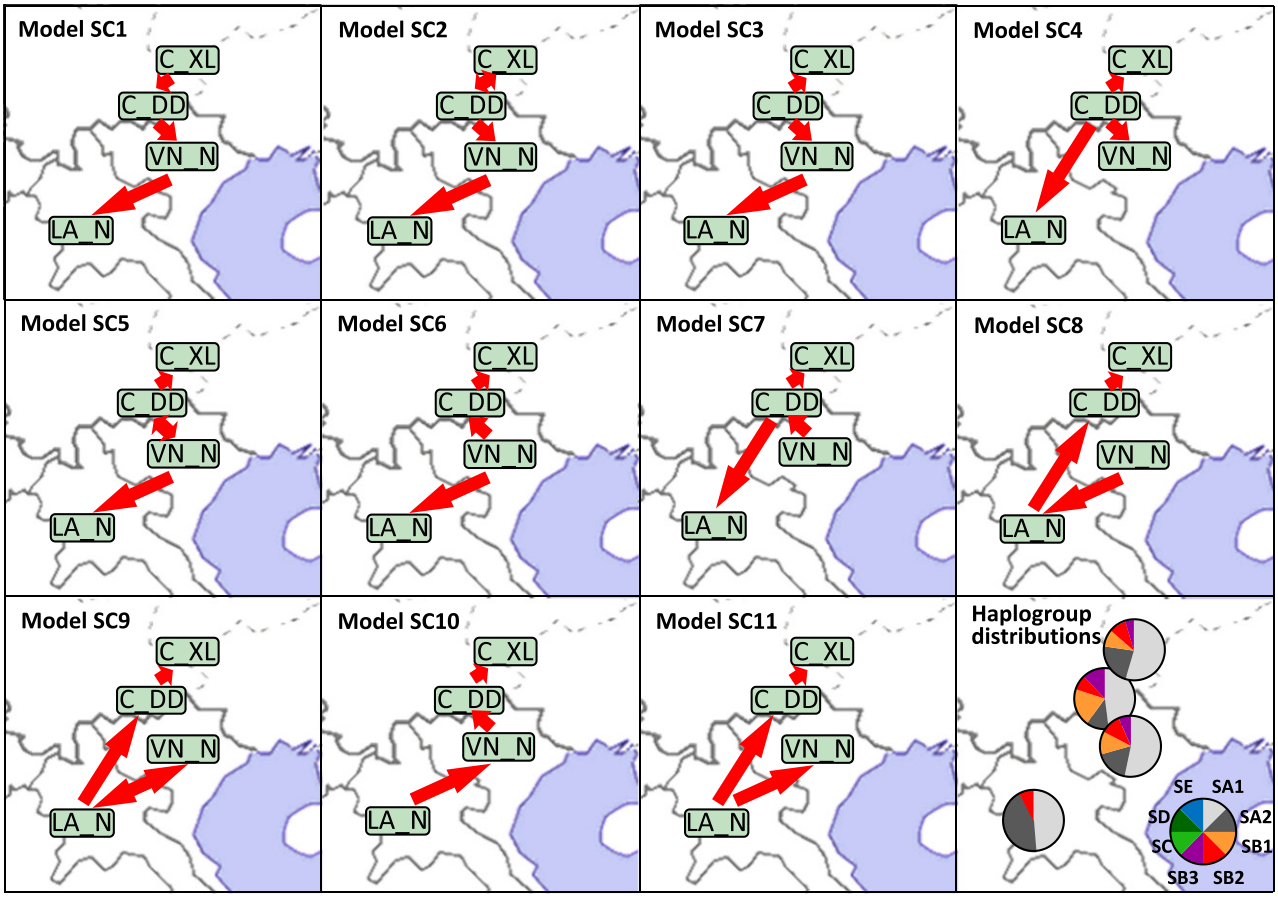

Fig. 11 Models of gene flow examined using the program MIGRATE-N, accounting for four populations on both sides of the south China border. MtDNA haplogroup distributions are indicated.

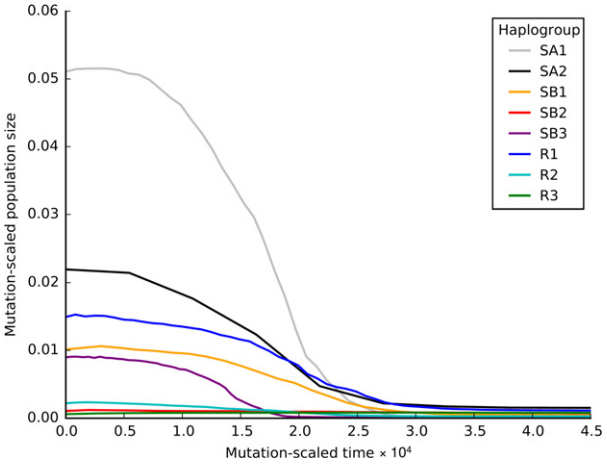

Fig. 12 Skyline plot for the swamp and river buffalo haplogroups.

Diversity of maternal lineages

The mtDNA control region and *Cytochrome b* gene are both widely used for phylogenetic and population genetic studies. The water buffalo phylogenies generated by these two fragments are entirely consistent (Fig. S3, Supporting information). However, only the high variability of the control region resolves the A1 and A2 and the B1, B2 and B3 haplogroups, respectively, that are essential for the detection of geographic trends. On the other hand, the high mutation rate in the control region leads to homoplasy and a saturation of the divergence (Alter & Palumbi 2009; Soubrier *et al.* 2012), which interfere with modelling approaches (Gerbault *et al.* 2014).

In our data, we observed relatively high values for haplotype diversity (*H*) of the control region and of the two loci combined of river buffaloes (Table 2) and an irregular pattern of *H* values across Chinese swamp buffalo populations (Table S6, Supporting information, Fig. 6), in contrast to the patterns based on *Cytochrome b* variability. We also identified and analysed Y-chromosomal haplotype variation. The strong geographic differentiation (Fig. 8) confirms the power of paternal lineages for reconstruction of population history (Groeneveld *et al.* 2010).

Separate trees of *Cytochrome b* and the control region incorporating published sequences from several sources (Fig. S3, Supporting information) showed an excellent consistency of the two mtDNA loci and also of our sequences with those previously published. All previous swamp sequences belong to the A and B haplogroups. Our sampling of river buffaloes in Bangladesh and southwest China covers only a small part of their broad distribution. However, we found that previously published sequences (Kierstein *et al.* 2004; Finlay *et al.* 2007; Ramadan & El-Hefnawi 2008; Lei *et al.* 2011; Yue *et al.* 2013), which together cover a large part of the geographic distribution of the river buffalo, did not define additional haplogroups (Fig. S3, Supporting information) and that all haplogroups are present throughout the distribution from China to the Mediterranean area. Remarkably, river mtDNA from southwest Asia, Egypt, Europe and South America all belong to the three haplogroups covered by our limited sampling of river buffalo. Thus, we can validly compare the genetic constitutions of river and swamp buffalo.

First, the branch lengths of the phylogenetic trees (Figs 3, S2 and S3, Supporting information) and the diversity parameters (Table 2) show that swamp buffalo harbour a higher sequence diversity than river buffalo. This is confirmed by a direct comparison of the Y-chromosomal variation as markers of the paternal lineages and suggests that the domestication of swamp buffalo involved a wider sampling of the wild ancestor than was the case for river buffalo.

Second, the intercontinental dispersal of the three haplogroups (Fig. S3, Supporting information, Nagarajan *et al.* 2015) indicates a phylogeographic structure of the maternal lineages of river buffalo that is weaker than observed for cattle (Lenstra *et al.* 2014), sheep (Cai *et al.* 2011), goat (Naderi *et al.* 2007) and even horse (Vilà *et al.* 2001; Jansen *et al.* 2002; Cieslak *et al.* 2010). We hypothesize that the westward migration of river buffalo from the Indian subcontinent through Mesopotamia and on to Europe was a more gradual expansion than for other livestock species and was without substantial population bottlenecks. This is confirmed by a gradual and modest decrease in the microsatellite heterozygosity from 0.71 to 0.78 in India (Kumar *et al.* 2006) to 0.58 to 0.68 in Italy (Moioli *et al.* 2001; Elbeltagy *et al.* 2008). In contrast, the distribution of the swamp buffalo mtDNA and Y-chromosomal haplogroups reveals a remarkable phylogeographic differentiation, which is further discussed below.

We have investigated the mtDNA diversity pattern of swamp buffalo on the basis of a broad representation of Chinese, Indochinese and Bangladeshi populations. Yue *et al.* (2013) studied a smaller data set covering Chinese populations only, which were proposed to have a weak phylogeographic structure. In agreement with Yue *et al.* (2013) and other previous studies (Lei *et al.* 2007), we found a major haplogroup A and a minor haplogroup B. However, a differentiation of two A and three B haplogroups, each having one major haplotype (Table 3), allowed us to detect the appreciable geographic differentiation mentioned above. This finding is entirely consistent with previous studies (Yue *et al.* 2013; Fig. S5, Supporting information) when their data were reanalysed on the basis of the haplogroups defined here. In this data set, the highest mtDNA diversity was found in Guangxi (GX, Fig. S5, Supporting information), which is close to the relatively diverse south Chinese populations from Xilin (XL) and Diandongnan (DD) and Vietnam North (VN-N) (Fig. 5). In southwest Chinese, Thai and Bangladeshi samples originating west of the Mekong River, we found a hitherto unidentified third haplogroup C and in other Thai samples haplogroups D and E.

Thus, the narrower ranges of our mtDNA haplogroups indicate a clear matrilocality of swamp buffalo (Fig. 5). This is remarkable not only because of the contrast with river buffalo, but also because of the absence of breed formation or any phenotypic differences among animals from different regions.

The geographic differentiation of mtDNA in the Chinese swamp buffalo is strong also in comparison with other livestock species in China, such as sheep (Zhao *et al.* 2013), pigs (Huo *et al.* 2016) and horses (Yue *et al.* 2012), but is comparable to the situation in goats (Zhao *et al.* 2014). For Chinese cattle, the strong phylogeographic structure (Lenstra *et al.* 2014) reflects the presence of both taurine and zebu, cross-fertile species that are as divergent as swamp and river buffalo (Yindee *et al.* 2010).

Maternal gene flow

MtDNA diversity patterns often reveal historic migration routes, which may date back to the dispersal of the first domesticates. Genetic variability is expected to decrease with increasing geographical distance from the centre of domestication (Groeneveld *et al.* 2010), unless there is introgression from the wild ancestor species outside the domestication site (Larson & Fuller 2014). Qualitatively, the absence of a haplogroup in a given region that is still present in neighbouring regions indicates a founder effect during or after gene flow from the neighbouring regions. Thus, we propose to explain the observed mtDNA haplotype distributions (Fig. 5) and the most likely gene-flow models (Figs 9–11, Tables S10–S13, Supporting information) by a series of gene flows:

1 starting in the region spanning the China/Indochina border;
2 then to the northeast via the coastal regions;
3 simultaneously to and along the Yangtze river, downstream as well as upstream to populations C_YJ, C_DC, C_DA and C_ER;
4 to the south across the Mekong and further to the south both east and west of the Mekong;
5 from Thailand and/or Myanmar to C_DH, C_ BL and Bangladesh.The occurrence of the rare C, D and E haplotypes and the high diversity west of the Mekong may then be explained by introgressive capture of wild cows (Larson & Fuller 2014). This scenario is compatible with quantitative clines of diversity (Fig. 6), with the haplotype diversity of *Cytochrome b* generating the most regular pattern. Mismatch distributions (Fig. S4, Supporting information), especially that of the SB3 haplogroup, indicate in agreement with Finlay *et al.* (2007), recent population expansions after the colonization of new regions. Estimation of female domestication times derived from mtDNA skyline plots (Table S14, Supporting information) tends to be shorter than indicated by the archaeological record, but are subject to a large uncertainty (see also Grant 2015). More reliable are the relative values for the different haplogroups. Thus, the shorter time estimates for haplogroups SB3 (swamp buffalo) and R2 (river buffalo) indicate a more recent expansion in agreement with lower numbers of mismatches (Fig. S4, Supporting information) and for SB3 also with a narrow post-domestic distribution. Remarkably, mismatch analysis suggested a recent expansion also for the major haplogroup A1, which is not incompatible with an earlier initial expansion detected by the skyline analysis.

Y-chromosomal diversity

In spite of clear perspectives for reconstruction of population history (Groeneveld *et al.* 2010; Yindee *et al.* 2010; Edwards *et al.* 2011), there have been few studies of Y-chromosomal variation in livestock species. Following up on the studies of Yindee *et al.* (2010), we identified several Y-chromosomal SNPs. Remarkably, the Y-chromosomal diversity of swamp buffalo is higher than that of other species, as we identified 11 haplotypes by sequencing of Y-chromosomal regions (*SRY*, *ZFY* and *DBY*), clearly more than found in the homologous sequences from other species: two haplotypes in river buffalo (this work), cattle (Edwards *et al.* 2011) or sheep (Meadows & Kijas 2009) and three in goats (Pereira *et al.* 2009). This appreciable diversity of paternal lineages does not indicate a horse-like scenario of many more female than male animals having been domesticated (Lippold *et al.* 2011; Wallner *et al.* 2013).

Analysis of paternal lineages also indicates that only the southern Chinese populations harbour all major haplotypes (Fig. 8). This supports the aforementioned scenario of a domestication that started in south China–north Indochina and continued west of the Mekong by incorporating wild females into the domestic population. Although sample sizes are small (Table S8, Supporting information), the high frequency of YS1 in the lower Yangtze River populations (Region 1), in contrast to its absence in some populations further to the south (Fig. 8), is likely due to importation of bulls from southern China, possibly promoted by the development of agriculture since the founding of the People's Republic of China in 1949 (Qiu 1984).

River buffalo Y haplotypes also were found in some of these lower Yangtze River populations. Murrah bulls were imported from India in 1957, Nili Ravi bulls from Pakistan in 1974 and Mediterranean bulls in 1996. These bulls or their descendants were used for cross-breeding programmes (China National Commission of Animal Genetics Resources 2011). These programmes were soon abandoned, but presumably left their traces in these populations. This, as well as river buffalo introgression in southwest China (C_BL, C_ER and C_GX) (China National Commission of Animal Genetics Resources 2011; Lei *et al.* 2011; Yue *et al.* 2013), has to be taken into account when comparing autosomal diversity patterns by high-density SNP genotypes or genomic sequences that are now becoming available.

Recent studies have concluded that there was a single domestication of rice 8200–13 500 years BP (Molina *et al.* 2011) that occurred in the middle region of the Pearl River valley in Guangxi Province (Huang *et al.* 2012). This province and surrounding regions also harbour the highest diversity of maternal and paternal lineages of the swamp buffalo (Figs 5, 8 and S5, Supporting information; Yue *et al.* 2013). It therefore seems plausible that the swamp buffalo domestication occurred in association with the spread of rice cultivation (Liu & Chen 2012).

However, definitive evidence for the time and site of domestication and the subsequent dispersal requires further studies, particularly for samples from south China and northern Indochina that include whole mtDNA sequences and autosomal markers.

Conclusions

We observed striking differences between swamp and river buffalo in the diversity patterns of the maternal and paternal lineages, with diversity much higher for the swamp type. For the swamp buffalo, both the

mtDNA and the Y-chromosomal diversity indicate a domestication of the swamp buffalo in the border region of south China and north Indochina. In contrast to the evidence for the domestication sites of other common livestock species, which is based on both archaeological and molecular evidence, our conclusions are based primarily on DNA analyses. It is also remarkable that the diversity on the DNA level is not accompanied by phenotypic differentiation among populations/breeds, as observed in most other livestock species. It will be interesting to see whether analysis of autosomal DNA reveals a geographic distribution of adaptive variation.

Acknowledgements

We thank Dongxiao Sun, Ying Yu, Shengli Zhang (China Agricultural University), Hui Deng (Peking University) and Jianlin Han (International Livestock Research Institute) for their valuable suggestions on the design of the study and interpretation of the results, Cecile Berthouly for help in selecting Vietnam samples, Anders Gonçalves da Silva for advice on migration models, Bertram Brenig and Gerold Kierstein for providing detailed information about their samples, Bruce Tier (AGBU, University of New England) and the Higher Performance Computing Center, Florida State University for provision of computing facilities and Peter Thomson and Isabelle Dupanloup for computing assistance. This work was supported by the National Natural Scientific Foundation of China (30901015, 31561143010), China Agricultural Research System (CARS-37) and the Program for Changjiang Scholars and Innovative Research in University (IRT1191). The contribution from Thailand was supported by the EU Asia Link Project (ASI/B7-301/98/679-035). Peter Beerli was supported by the US National Science Foundation (DEB-1145999). We appreciate the several suggestions of the reviewers.

References

Ajmone-Marsan P, Garcia JF, Lenstra JA (2010) On the origin of cattle: how aurochs became domestic and colonized the world. *Evolutionary Anthropology*, **19**, 148–157.

Alter SE, Palumbi SR (2009) Comparing evolutionary patterns and variability in the mitochondrial control region and cytochrome B in three species of baleen whales. *Journal of Molecular Evolution*, **68**, 97–111.

Amano T, Namikawa T, Suzuki S (1980) Genetic differences between swamp and river buffaloes in the electrophoretic variations of albumin and transferrin. *Proceedings of the Japan Academy. Series B, Physical and biological Sciences*, **56**, 463–468.

Bandelt HJ, Forster P, Röhl A (1999) Median–joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Barker JSF, Tan SG, Selvaraj OS, Mukherjee TK (1997a) Genetic variation within and relationships among populations of Asian water buffalo (*Bubalus bubalis)*. *Animal Genetics*, **28**, 1–13.

Barker JSF, Moore SS, Hetzel DJS, Evans D, Byrne K, Tan SG (1997b) Genetic diversity of Asian water buffalo (*Bubalus bubalis*): microsatellite variation and a comparison with protein-coding loci. *Animal Genetics*, **28**, 103–115.

Beerli P (2006) Comparison of Bayesian and maximum likelihood inference of population genetic parameters. *Bioinformatics*, **22**, 341–345.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, **185**, 313–326.

Beja-Pereira A, Caramelli D, Lalueza-Fox C *et al.* (2006) The origin of European cattle: evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences, USA*, **103**, 8113–8118.

Bollongino R, Burger J, Powell A, Mashkour M, Vigne J-D, Thomas MG (2012) Modern taurine cattle descended from small number of near-eastern founders. *Molecular Biology and Evolution*, **29**, 2101–2104.

Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Bradley DG, MacHugh DE, Cunningham P, Loftus RT (1996) Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences, USA*, **93**, 5131–5135.

Bruford MW, Bradley DG, Luikart G (2003) DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, **4**, 900–910.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Verlag, New York.

Cai D, Tang Z, Yu H *et al.* (2011) Early history of Chinese domestic sheep indicated by ancient DNA analysis of Bronze Age individuals. *Journal of Archaeological Science*, **38**, 896–902.

Cañón J, García D, García-Atance MA *et al.* (2006) Geographical partitioning of goat diversity in Europe and the Middle East. *Animal Genetics*, **37**, 327–334.

Chen S, Lin B-Z, Baig M *et al.* (2010) Zebu cattle are an exclusive legacy of the south Asia Neolithic. *Molecular Biology and Evolution*, **27**, 1–6.

Chessa B, Pereira F, Arnaud F *et al.* (2009) Revealing the history of sheep domestication using retrovirus integrations. *Science*, **324**, 532–536.

China National Commission of Animal Genetics Resources (2011) *Animal Genetic Resources in China: Bovines*. pp. 10–11. China Agriculture Press, Beijing. (in Chinese).

Cieslak M, Pruvost M, Benecke N *et al.* (2010) Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS ONE*, **5**, e15311.

Cockrill WR (1984) Water buffalo. In: *Evolution of Domesticated Animals*(ed. Mason IL), pp. 52–63. Longman, London.

Corander J, Sirén J, Arjas E (2008) Bayesian spatial modelling of genetic population structure. *Computational Statistics*, **23**, 111–129.

Decker JE, McKay SD, Rolf MM *et al.* (2014) Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genetics*, **10**, e1004254.

Diamond J (1997) *Guns, Germs and Steel*. Jonathan Cape, London.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.

Edwards CJ, Ginja C, Kantanen J *et al.* (2011) Dual origins of dairy cattle farming – evidence from a comprehensive survey of European Y-chromosomal variation. *PLoS ONE*, **6**, e15922.

Elbeltagy AR, Galal S, Abdelsalam AZ, El Keraby FE, Blasi M, Mohamed MM (2008) Biodiversity in Mediterranean buffalo using two microsatellite multiplexes. *Livestock Science*, **114**, 341–346.

Epstein H (1969) *Domestic Animals of China*. Commonwealth Agricultural Bureaux, Farnham Royal.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Fang M, Andersson L (2006) Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proceedings of the Royal Society B*, **273**, 1803–1810.

FAO (2000) Wild relatives of domestic livestock and some suggestions for new domesticants. In: *World Watch List for Domestic Animal Diversity*, 3rd Edn. (ed. Scherf B), pp. 647–716. FAO, Rome.

FAO (2007) Origin and history of livestock diversity. In: *The State of the World's Animal Genetic Resources for Food and Agriculture* (eds Rischkowsky B, Pilling D), pp. 5–22. FAO, Rome.

Finlay EK, Gaillard C, Vahidi SMF *et al.* (2007) Bayesian inference of population expansions in domestic bovines. *Biology Letters*, **3**, 449–452.

Fischer H, Ulbrich F (1968) Chromosomes of the Murrah buffalo and its crossbreds with the Asiatic swamp buffalo (*Bubalus bubalis*). *Zeitschrift fürTierzüchtung und Züchtungsbiologie*, **84**, 110–114.

Flamand JRB, Vankan D, Gairhe KP, Duong H, Barker JSF (2003) Genetic identification of wild Asian water buffalo in Nepal. *Animal Conservation*, **6**, 265–270.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Fu Y-X, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.

Gerbault P, Allaby RG, Boivin N *et al.* (2014) Storytelling and story testing in domestication. *Proceedings of the National Academy of Sciences, USA*, **111**, 6159–6164.

Giuffra E, Kijas JMH, Amarger V, Carlborg Ö, Jeon J-T, Andersson L (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, **154**, 1785–1791.

Grant WS (2015) Problems and cautions with sequence mismatch analysis and Bayesian skyline plots to infer historical demography. *Journal of Heredity*, **106**, 333–346.

Groeneveld LF, Lenstra JA, Eding H *et al.* (2010) Genetic diversity in farm animals – a review. *Animal Genetics*, **41**(Suppl. 1), 6–31.

Higham C (2002) *Early Cultures of Mainland Southeast Asia*. River Books, Bangkok.

Huang X, Kurata N, Wei X *et al.* (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature*, **490**, 497–503.

Huo JH, Wei QP, Wan MC *et al.* (2016) Population phylogenomic analysis and origin of mitochondrial DNA in Chinese domestic pig. *Mitochondrial DNA*, **27**, 892–895.

Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, **14**, 68–73.

Jansen T, Forster P, Levine MA *et al.* (2002) Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences, USA*, **99**, 10905–10910.

Kierstein G, Vallinoto M, Silva A, Schneider MP, 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**, 308–324.

Kijas JW, Lenstra JA, Hayes B *et al.* (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, **10**, e1001258.

Kumar S, Gupta J, Kumar N *et al.* (2006) Genetic variation and relationships among eight Indian riverine buffalo breeds. *Molecular Ecology*, **15**, 593–600.

Kumar S, Nagarajan M, Sandhu JS, Kumar N, Behl V, Nishanth G (2007a) Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. *Animal Genetics*, **38**, 227–232.

Kumar S, Nagarajan M, Sandhu JS, Kumar N, Behl V (2007b) Phylogeography and domestication of Indian river buffalo. *BMC Evolutionary Biology*, **7**, 186–193.

Larson G, Fuller DQ (2014) The evolution of animal domestication. *Annual Review of Ecology, Evolution and Systematics*, **45**, 115–136.

Lau CH, Drinkwater RD, Yusoff K, Tan SG, Hetzel DJS, Barker JSF (1998) Genetic diversity of Asian water buffalo (*Bubalus bubalis*): mitochondrial DNA D-loop and cytochrome b sequence variation. *Animal Genetics*, **29**, 253–264.

Lei CZ, Zhang W, Chen H *et al.* (2007) Independent maternal origin of Chinese swamp buffalo (*Bubalus bubalis*). *Animal Genetics*, **38**, 97–102.

Lei CZ, Zhang CM, Weining S *et al.* (2011) Genetic diversity of *mitochondrial cytochrome b* gene in Chinese native buffalo. *Animal Genetics*, **42**, 432–436.

Lenstra JA, Groeneveld LF, Eding H *et al.* (2012) Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Animal Genetics*, **43**, 483–502.

Lenstra JA, Ajmone-Marsan P, Beja-Pereira A *et al.* (2014) Meta-analysis of mitochondrial DNA reveals several population bottlenecks during worldwide migrations of cattle. *Diversity*, **6**, 178–187.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lippold S, Knapp M, Kuznetsova T *et al.* (2011) Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nature Communications*, **2**, 450.

Liu L, Chen X (2012) *The Archaeology of China: From the Late Paleolithic to the Early Bronze Age*. Cambridge University Press, Cambridge.

Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P (1994) Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences, USA*, **91**, 2757–2761.

Luikart G, Gielly L, Excoffier L, Vigne J-D, Bouvet J, Taberlet P (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences, USA*, **98**, 5927–5932.

Macgregor R (1941) The domestic buffalo. *The Veterinary Record*, **58**, 443–450.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Meadows JRS, Kijas JW (2009) Re-sequencing regions of the ovine Y chromosome in domestic and wild sheep reveals novel paternal haplotypes. *Animal Genetics*, **40**, 119–123.

Meadows JRS, Cemal I, Karaca O, Gootwine E, Kijas JW (2007) Five ovine mitochondrial lineages identified from sheep breeds of the near east. *Genetics*, **175**, 1371–1379.

Moioli B, Georgoudis A, Napolitano F *et al.* (2001) Genetic diversity between Italian, Greek and Egyptian buffalo populations. *Livestock Production Science*, **70**, 203–211.

Molina A, Sikora M, Garud N *et al.* (2011) Molecular evidence for a single evolutionary origin of domesticated rice. *Proceedings of the National Academy of Sciences, USA*, **108**, 8351–8356.

Naderi S, Rezaei H-R, Taberlet P *et al.* (2007) Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS ONE*, **2**, e1012.

Naderi S, Rezaei H-R, Pompanon F *et al.* (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences, USA*, **105**, 17659–17664.

Nagarajan M, Nimisha K, Kumar S (2015) Mitochondrial DNA variability of domestic river buffalo (Bubalus bubalis) populations: genetic evidence for domestication of river buffalo in Indian subcontinent. *Genome Biology and Evolution*, **7**, 1252–1259.

Nijman IJ, Van Boxtel DCJ, Van Cann LM *et al.* (2008) Phylogeny of Y-chromosomes from bovine species. *Cladistics*, **24**, 723–726.

Nomura K, Yonezawa T, Mano S *et al.* (2013) Domestication process of the goat revealed by an analysis of the nearly complete mitochondrial protein-encoding genes. *PLoS ONE*, **8**, e67775.

Patel AK, Meadow RH (1998) The exploitation of wild and domestic water buffalo in prehistoric northwestern south Asia. In: *Archaeozoology of the Near East III: Proceedings of the Third International Symposium on the Archaeozoology of Southwestern Asia and Adjacent Areas* (eds Buitenhuis H, Bartosiewicz L, Choyke AM), **18**, pp. 180–199. ARC Publicaties, Groningen.

Pereira F, Queirós S, Gusmão L *et al.* (2009) Tracing the history of goat pastoralism: new clues from mitochondrial and Y chromosome DNA in North Africa. *Molecular Biology and Evolution*, **26**, 2765–2773.

Qiu H (1984) Jianguo Yilai Woguo Yangniuye Fazhan Gaikuang (Overview of the cattle industry in China since the founding of the People's Republic of China). *Xumu Shouyi Zazhi*, **4**, 27–34. (In Chinese).

Ramadan HA, El-Hefnawi MM (2008) Phylogenetic analysis and comparison between cow and buffalo (including Egyptian buffaloes) mitochondrial displacement-loop regions. *Mitochondrial DNA*, **19**, 401–410.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Shapiro B, Drummond AJ, Rambaut A *et al.* (2004) Rise and fall of the Beringian steppe bison. *Science*, **306**, 1561–1565.

Soubrier J, Steel M, Lee MSY *et al.* (2012) The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Molecular Biology and Evolution*, **29**, 3345–3358.

Swofford DL (2003) *PAUP: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4*. Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tapio M, Marzanov N, Ozerov M *et al.* (2006) Sheep mitochondrial DNA variation in European, Caucasian, and central Asian areas. *Molecular Biology and Evolution*, **23**, 1776–1783.

Ulbrich F, Fischer H (1967) The chromosomes of the Asiatic buffalo (*Bubalus bubalis*) and the African buffalo (*Cyncerus caffer*). *Zeitschrift fürTierzüchtung und Züchtungsbiologie*, **83**, 219–223.

Vilà C, Leonard JA, Götherström A *et al.* (2001) Widespread origins of domestic horse lineages. *Science*, **291**, 474–477.

Wallner B, Vogl C, Shukla P, Burgstaller JP, Druml T, Brem G (2013) Identification of genetic variation on the horse Y chromosome and the tracing of male founder lineages in modern breeds. *PLoS ONE*, **8**, e60015.

Wang J, Zhang J (2011) Shengshuiniu de jiayang yesheng shuxing chubu yanjiu (Preliminary study on the domesticated or wild character of *Bubalus mephistopheles*). *Nanfang Wenwu*, **3**, 134–139. (In Chinese).

Warmuth V, Eriksson A, Bower MA *et al.* (2012) Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proceedings of the National Academy of Sciences, USA*, **109**, 8202–8206.

Wiener P, Wilkinson S (2011) Deciphering the genetic basis of animal domestication. *Proceedings of the Royal Society B*, **278**, 3161–3170.

Yindee M (2010) *Reproduction and genetic diversity of the swamp buffalo*. Acad. Thesis, Utrecht University, Utrecht. ISBN 978-90-393-5399-8. (Chapter 5 – High diversity of the mitochondrial DNA of the Thai swamp buffalo indicates a domestication in Southeast Asia, pp. 63-76). http://igitur archive.library.uu.nl/dissertations/2011-0121-200244/Yindee.pdf.

Yindee M, Vlamings BH, Wajjwalku W *et al.* (2010) Y-chromosomal variation confirms independent domestications of swamp and river buffalo. *Animal Genetics*, **41**, 433–435.

Yue X-P, Qin F, Campana MG *et al.* (2012) Characterization of *cytochrome b* diversity in Chinese domestic horses. *Animal Genetics*, **43**, 624–626.

Yue X-P, Li R, Xie W-M *et al.* (2013) Phylogeography and domestication of Chinese swamp buffalo. *PLoS ONE*, **8**, e.56552.

Zhang Y, Sun D, Yu Y, Zhang Y (2006) A Y-linked SNP in *SRY* gene differentiates Chinese indigenous swamp buffalo and

introduced river buffalo. *Asian-Australasian Journal of Animal Science*, **19**, 1240–1244.
Zhang Y, Vankan D, Zhang Y, Barker JSF (2011) Genetic differentiation of water buffalo (*Bubalus bubalis*) populations in China, Nepal and south-east Asia: inferences on the region of domestication of the swamp buffalo. *Animal Genetics*, **42**, 366–377.
Zhao E, Yu Q, Zhang N, Kong D, Zhao Y (2013) Mitochondrial DNA diversity and the origin of Chinese indigenous sheep. *Tropical Animal Health and Production*, **45**, 1715–1722.
Zhao Y, Zhao R, Zhao Z, Xu H, Zhao E, Zhang J (2014) Genetic diversity and molecular phylogeography of Chinese domestic goats by large-scale mitochondrial DNA analysis. *Molecular Biology Reports*, **41**, 3695–3704.

Y.Z., M.Y., Y.-T.J., B.C., Y.Z., J.A.L. and J.S.F.B. conceived the experiments. Y.Z., M.Y., Y.-T.J., S.Y., M.O.F., Y.W., V.C.C., L.D.P., B.B., B.Y., X.L., Z.C., D.V., W.M., N.K., S.D. and Y.Z. contributed samples. Y.Z., Y.L., M.Y., K.-Y.L., Q.L. and Y.Z. conducted the experiments. Y.Z., P.B., J.A.L. and J.S.F.B. analysed the data. Y.Z., J.A.L., P.B. and J.S.F.B. wrote the manuscript.

Data accessibility

DNA sequences: GenBank accessions KR007969–KR010168, GQ259327–GQ259332, KT186376–KT186427.

Data analysis input files: Dryad doi: http://dx.doi.org/10.5061/dryad.16mp4.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1. Sequence variation of water buffalo control region and *Cytochrome b*.

Fig. S2. Maximum likelihood tree of water buffalo mtDNA haplotypes.

Fig. S3. Neighbor-joining trees of *Cytochrome b* and the control region haplotypes from this and previous work.

Fig. S4. Number of mismatches in pairwise comparisons of sequences from the same haplogroup.

Fig. S5. Haplogroup distributions of swamp buffaloes sampled by Yue *et al.* (2013).

Supplement to Table 1. Coordinates for populations sampled from several localities.

Table S1. Primers for amplification and sequencing of mitochondrial control region and *Cytochrome b* gene and Y-chromosomal *ZFY*, *SRY* and *DBY* segments of water buffalo.

Table S2. AMOVA analysis of mtDNA and Y-chromosomal DNA of swamp buffalo populations.

Table S3. Populations, sample sizes and frequencies of mtDNA haplogroups per population.

Table S4. Tajima *D* and Fu *Fs* statistics for river and swamp buffalo mtDNA haplogroups.

Table S5. Sequence variation of swamp type haplogroups within regional groups.

Table S6. MtDNA sequence diversity for swamp and river populations.

Table S7. Y haplotypes defined by SNPs in *ZFY1, SRY* and *DBY*.

Table S8. Y-chromosome sample sizes and haplotype counts.

Table S9. MtDNA and Y-chromosomal sequence diversity for the 17 groups of swamp populations and the mtDNA diversity for the 7 regions (Fig. 2).

Table S10. Marginal log-likelihoods and model probabilities for seven migration models (Fig. 9) among three main areas.

Table S11. Marginal log-likelihoods and model probabilities for 17 migration models (Fig. 10) among seven regions.

Table S12. Marginal log-likelihoods and model probabilities for seven migration models (Fig. 10) among seven regions.

Table S13. Marginal log-likelihoods and model probabilities for 11 migration models (Fig. 11) among four populations from South China and North Indochina.

Table S14. Population expansion time estimates on the basis of Bayesian skyline analysis.