Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 30 (2004) 308–324
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Analysis of mitochondrial D-loop region casts new light on domestic water buffalo (*Bubalus bubalis*) phylogeny☆

Gerold Kierstein,[a,b] Marcelo Vallinoto,[c] Artur Silva,[c] Maria Paula Schneider,[c] Leopoldo Iannuzzi,[d] and Bertram Brenig[a,*]

[a] *Institute of Veterinary Medicine, Georg-August University, Groner Landstrasse 2, 37073 Göttingen, Germany*
[b] *International Livestock Research Institute, P.O. Box 30709 Nairobi, Kenya*
[c] *Centro de Ciencias Biológicas, Departamento de Genética, Laboratorio de Polimorfismo de DNA, Universidade Federal do Pará, Belém, Brazil*
[d] *Consiglio Nazionale delle Ricerche (CNR), Istituto per il Sistema Produzione Animale in Ambiente Mediterraneo (ISPAAM), Naples, Italy*

Received 15 May 2002

Abstract

The phylogeny of water buffaloes (*Bubalus bubalis*) is still a matter of discussion, especially if the two types of domestic water buffalo (swamp and river) derived from different domestication events or if they are products of human selection. To obtain more insight, we analyzed the entire mitochondrial D-loop region of 80 water buffaloes of four different breeds, i.e., 19 swamp buffaloes (Carabao) and 61 river buffaloes (Murrah, Jafarabadi, and Mediterranean), sampled in Brazil and Italy. We detected 36 mitochondrial haplotypes with 128 polymorphic sites. Pooled with published data of South-East Asian and Australian water buffaloes and based on comprehensive median-joining network and population demography analyses we show evidence that both river and swamp buffaloes decent from one domestication event, probably in the Indian subcontinent. However, the today swamp buffaloes have an unravelled mitochondrial history, which can be explained by introgression of wild water buffalo mtDNA into domestic stocks. We are also discussing indications for an independent domestication of buffaloes in China.

Keywords: Domestication; *Bubalus*; Water buffalo; Mitochondrial DNA; D-loop region; Median-joining network

1. Introduction

To date, little is known about the genetics of the water buffalo (*Bubalus bubalis* L.). This might partly be due to the lower performance of buffaloes with respect to economic important traits when compared to cattle, limiting the interests in buffalo genome analysis in the developed countries. However, the domestic water buffalo holds a great economic potential in the developing countries. Here the water buffalo is admired as multipurpose animal for diary, meat, and drought. The stock of domestic water buffalo is estimated to be around 130 million, which is around 1/9 of the total worldwide cattle population. Nevertheless, there are more people in the world depending on domestic water buffalo than on any other domesticated species (FAO and UNEP, 2000). To develop rational breeding programs knowledge of extent and pattern of genetic variability within a breed or a population is essential. Furthermore, it is a prerequisite to the conservation of genetic resources.

Within the family of Bovidae, the tribe of Bovini consists of the genera *Bos*, *Bison*, *Pseudoryx*, *Bubalus*, and *Syncerus* (Hassanin and Douzery, 1999a). The latter two genera belong to the group of buffaloes, the Asian and African buffaloes. These are different lineages, separated from the other bovines during the Pliocene, demonstrated by archaeological findings of *Syncerus* remains in South Africa (Savage and Russell, 1983) or even in the end of the late Miocene demonstrated by molecular evolution of the bovid cytochrome *b* gene (Hassanin and Douzery, 1999b). During the Pleistocene, *Bubalus* was distributed from southern Asia to Europe. With increasing dry climate the area of distribution decreased to

☆ GenBank Accession Nos. AF197196–AF197223 and AY195589–AY195599.

* Corresponding author. Fax: +49-551-393392.
E-mail address: bbrenig@gwdg.de (B. Brenig).

1055-7903/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S1055-7903(03)00221-5

India, Indonesia, and parts of Southeast Asia (Cockrill and Mahadevan, 1974; Nachtsheim and Stengel, 1977). The genus *Bubalus* is divided into the subgenera Bubalus and Anoa with four living wild species (Tanaka et al., 1996). The group of Anoa encompasses the Lowland anoa (*B. depressicornis*) and the Mountain anoa (*B. quarlesi*) on Sulawesi, Indonesia, and the Tamaraw (*B. mindorensis*) on Mindoro, Philippines. The wild water buffalo (*B. bubalis*), inhabiting India, is thought to be the founder of the today domestic buffaloes, which are subdivided into the swamp ($2n = 48$) and river ($2n = 50$) buffaloes using karyotyping (Fischer and Ulbrich, 1968; Iannuzzi, 1994; Ulbrich and Fischer, 1967) as well as morphological and ethological criteria (for review see Cockrill, 1981; Cockrill and Mahadevan, 1974).

The domestication of buffaloes most likely took place in the civilization of the Indus, the Yangtze, and the Euphrates and Tigris in the third millennium BC (Cockrill, 1981; Nachtsheim and Stengel, 1977) and/or in China during the fifth millennium BC (Chen and Li, 1989). Buffaloes were introduced to Italy from central Europe in the sixth century or by the Bey of Tunis in the seventh century at the time of the Arab conquest (Salerno, 1974). Importation of water buffaloes to Africa, Australia, and South America took place only recently. The common understanding about water buffaloes is the division into two types, with at least 18 well-defined river buffalo, whereas the swamp buffalo solitary exists with one breed (Cockrill, 1981) sometimes referred to as Carabao. Cockrill (1974) describes in his introduction the Malay origin of the word Carabao, used in the Philippines to distinguish the local swamp buffalo from the introduced Indian river buffalo. He recommends not applying the word Carabao for any type of water buffalo. However, as at least a part of the Brazilian swamp buffaloes were imported from the Philippines still today the terminus Carabao is used in Brazil. Therefore, in this manuscript we are not strictly following this nomenclature recommended by W. Ross Cockrill, especially when referring to the Brazilian swamp buffalo population. Nevertheless, especially in the southeastern provinces of China a huge variety of genetic resources of swamp buffaloes have been described in terms of attitudes, adaptation to environment and distribution (Chunxi and Zhongquan, 2001). The authors referred to 18 different breeds of swamp buffaloes, evolved separately in each area, but no genetic study has been conducted so far.

Despite a variety of investigations on water buffalo's molecular phylogeny the exact phylogenetic relationship between the two types of water buffaloes is still discussed. The estimations of divergence of river and swamp buffaloes range from 10,000 to 1.7 million years (Amano et al., 1994; Barker et al., 1997a,b; Lau et al., 1998; Ritz et al., 2000; Tanaka et al., 1996, 1995). These extreme differences depend on the variety of methods (RFLP, sequencing of mitochondrial coding and non-coding regions, microsatellite analysis) and on the sample size introduced to the analysis. The determination of the divergence of buffaloes is of interest, because the current estimations of the more recent divergence (Lau et al., 1998) coincide with the presumed domestication time of buffaloes. Hence, buffaloes either diverged into the river and swamp buffaloes after domestication, thus by human selection, or the two types were domesticated independently.

To shade more light on this subject we started a comprehensive analysis on water buffalo phylogeny using the sequence information of the complete mitochondrial D-loop region. Due to the maternal inheritance and the lack of recombination of mitochondrial DNA, the mutations accumulated throughout a species' history trace the maternal genealogy. The usage of the mitochondrial cytochrome *b* gene has been proven as suitable marker to reconstruct the phylogeny within the family of Bovidae (Hassanin and Douzery, 1999b). Hence, the more variable part of the mitochondrial DNA (i.e., the D-loop region) allows the genealogy on population level to be captured in a fair amount of detail, as has been shown in other species like cattle, sheep, and deer to name only a few (Boyce et al., 1999; Cymbron et al., 1999; Douzery and Randi, 1997; Nagata et al., 1999; Troy et al., 2001).

However, on population level the treelike relationship (as implemented in many phylogenetic software) can be obscure due to the mode of inheritance. By nature, relationships between different species are hierarchically as a consequence of reproductive isolation and fixation of alleles, which eventually result in non-overlapping gene pools. In contrast, within a species relationships are not hierarchically as they are the result of reproduction (Posada and Crandall, 2001). Therefore, for the intraspecific analysis we focused on the median-joining (MJ) network algorithm (Bandelt et al., 1999), which is especially applicable to non-recombinant DNA sequences, like mitochondrial DNA. This approach, similar to that of Foulds et al. (1979), combines all minimum spanning trees to a single network. Using a Parsimony criterion, median vectors, i.e., the missing intermediates, are added to the network. The advantage of this method is that a network has the capability to reveal conflicts among different sites, whereas a bifurcated tree may represent an incorrect or unresolved representation of a relationship. We further employed a population demography analysis to support our findings obtained in the network analysis.

2. Materials and methods

2.1. Data collection

A total of 63 blood samples were collected from Murrah, Jafarabadi, and Carabao breeds from different locations in Brazil. A total of 17 samples (Mediterranean)

Table 1
Origin of blood and/or genomic DNA samples from different buffalo breeds

Origin	Breed	Sample	Animal No.
Fazenda Itaquí—Castanhal, Pará (Brazil)	Jafarabadi	Fresh blood	01, 02, 03, 04, 19, 23, 25, 27, 30, 32, 40, 44, 47, 49, 50, 51, 53, 55, 60, 66
	Murrah	Fresh blood	105, 110, 117
EMBRAPA/CPATU—Belém, Pará (Brazil)	Murrah	Fresh blood	143, 181, 182, 183
Fazenda São Luís—Marajó, Pará (Brazil)	Murrah	Fresh blood	196, 267, 304, 310, 314
	Carabao	Fresh blood	315, 320, 325, 327, 330, 335, 336, 341, 342, 344, 348, 349, 352, 359, 366, 370, 375, 377, 380
Fazenda Jari—Monte Dairado, Amapá (Brazil)	Murrah	Fresh blood	481, 497
Registro, São Paulo (Brazil)	Murrah	Fresh blood	IZ278, IZ192, IZ358, IZ244, IZ369, IZ143, IZ364, IZ287, IZ368, WB1649
ISPAAM (Italy)[a]	Mediterranean	Fresh blood	B05, C01, C02, C03, C04, C05, C07, C08
		DNA	D01, D02, D03, D04, D05, D06, D07, D09, D10

[a] Letters indicate that animals were raised on the same farm.

were collected in Italy (see also Fig. 3d). This dataset is thereafter referred as "Brazilian/Italian" samples. The origins and types of samples are summarized in Table 1. Samples were taken randomly and the breed was phenotypically determined.

For comparison with our data additional *Bubalus* mitochondrial D-loop sequences were obtained from GenBank (Accession. No. AF016397, Lau et al., 1998), thereafter referred as "South-East Asian/Australian" samples. For certain parts of the analyses, where a complete D-loop sequence (or a certain part) was required, the bovine sequences were employed (Accession No. NC_001567, Anderson et al., 1982).

2.2. Extraction of DNA from blood samples

Blood was freshly taken from the jugular vein (5 ml into 300 µl EDTA) and diluted to 12 ml with 1 × SSC buffer, mixed and centrifuged. The leukocyte pellet was resuspended in 12 ml 1 × SSC and centrifuged. This step was repeated several times until the supernatant was clear. The leukocyte pellet was resuspended with Na-acetate (0.2 M, pH 7.0) prior to adding 0.5 ml 10% SDS. The solution was gently mixed until it became clear. Total genomic DNA was extracted with an equal volume of phenol/chloroform. The aqueous phase was re-extracted with an equal volume of chloroform/isoamyl-alcohol (24/1). The DNA was precipitated by adding an equal volume of isopropanol, washed with ethanol, air-dried, and resolved in 1 ml TE buffer (10/1, pH 8.0).

The blood samples of the Italian buffaloes were delivered in 5 ml vacutainers. DNA was extracted by a simple modified salting out procedure as described elsewhere (Miller et al., 1988).

2.3. Amplification and sequencing of mtDNA

To amplify the mitochondrial D-loop region, primers (Table 2) were designed based on the known *Bos taurus* mtDNA sequence (Anderson et al., 1982), positioned in the conserved $tRNA^{Phe}$ and Cytb genes. For the DNA sequence analysis of the Carabao specimens four internal primers had to be designed additionally as all Carabao samples displayed a G-stretch, resulting in unreadable sequence data. Hence, we amplified the G-stretch region separately, resulting in a 188 bp PCR product, which overlapped with two further PCR fragments (159 and 716 bp) harbouring the borders of the 188 bp fragment and the conserved regions of the $tRNA^{Phe}$ and Cytb genes, respectively. For direct DNA sequencing all primers were tailed with universal M13 forward primer (5′-TGT AAA ACG ACG GCC AGT-3′) and universal M13 reverse primer (5′-CAG GAA ACA GCT ATG ACC-3′).

PCR was carried out in a total volume of 100 µl using 10 ng of isolated DNA, 100 pmol of each primer,

Table 2
Oligonucleotides used for DNA sequencing of the bubaline mitochondrial D-loop region

Name[a]	sequence
$tRNA^{Phe}$	5′-AGG CAT TTT CAG TGC CTT GC-3′
Cytb	5′-TAG TGC TAA TACCAA CGGCC-3′
G-stretch-uni	5′-CCA TCA ACA CAC CTG ACC-3′
G-stretch-rev	5′-GCG AGG ACG GAT TTG ACT-3′
D-loop-int-uni	5′-CCA TTC GGA GTA GTA GGG TC-3′
D-loop-int-rev	5′-CAT AAC ATT AAT GTA ATA AGG GC-3′

[a] The first two oligonucleotides anneal to the conserved regions outside the D-loop. The other four are internal oligonucleotides to the guanine-rich region of the Carabao D-loop region.

200 μM dNTPs, 2.5 units *Taq* polymerase, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, and 1.5 mM $MgCl_2$. Best results were obtained using the following reaction conditions on a RoboCycler (Stratagene, Heidelberg): 95 °C (1 min), 35 cycles of 94 °C (1 min), 56 °C (1 min), 72 °C (1 min), followed by a final step at 72 °C (4 min). PCR products were electrophoresed on 1% agarose gels and purified using the QIAEX II kit (Qiagen, Hilden) according to the supplier's instructions. Purified DNA was sequenced using the SEQUENASE Cycle Sequencing kit (Amersham Pharmacia, Freiburg). Reactions were applied to a 6% polyacrylamide gel and sequenced automatically (LI-COR, MWG Biotech, Ebersberg). Both DNA strands were checked for ambiguous bases and edited manually. The DNA sequences were aligned using the SEQUENCHER (3.0) software (Gene Codes Inc.) and the resulting consensus sequence for each animal was used for the phylogenetic analyses.

2.4. Phylogenetic analyses

Multiple sequence alignments were performed using the CLUSTALX (1.64b) program (Thompson et al., 1997). Alignments were optimized by realigning selected sequence ranges and finally by manual editing using MACCLADE 4.0 (Maddison and P., 2000).

Maximum likelihood (ML) (Felsenstein, 1981) reconstructions were performed using the quartet puzzling (QP) tree search in TREE-PUZZLE 5.0 (Schmidt et al., 2002). After removal of sites where insertion/deletions (indels) occurred, the substitution model of Tamura and Nei (1993) and an eight-category gamma distribution of the substitution rates across variable sites (Yang, 1996) were chosen. The robustness of internal branches was assessed by the QP reliability after 10,000 puzzling steps, where reliability percentages (RP) above 90% can be considered as very strong supported (Schmidt et al., 2002).

For median-joining (MJ) network analysis all variable characters of the complete alignment were entered into the program package NETWORK 3.1.1.1 (Bandelt et al., 1999). As the algorithm is based on the Hamming distance, deletions are treated as a fifth character by default. However, as a refinement one may weigh character changes. Two approaches were employed to deal with gaps: (i) gaps were treated as missing character states or (ii) treated as one evolutionary event by weighing adjacent gaps as zero. To cope with the different probability of transition/transversion occurrence and the different substitution rates in the variable region and in the conserved central domain of the mtDNA, characters outside the conserved central domain were weighed with 1 (transitions) or 2 (transversions). Within the conserved central domain transitions were weighed with 3 and transversions with 4. Alternatively, the data were analysed with all character states weighed equally. Bandelt and his colleagues (1999) specified a tolerance parameter ε, which widens the search for potential new median vectors when increasing the value of ε. The network calculations were run with different values for ε, however the best results were obtained with $\varepsilon = 0$. Frequencies of haplotypes were converted into proportional areas in the figures. The constructed networks were imported to a vector-based drawing software to redraw a sophisticated illustration.

2.5. Population analyses

The pooled Brazilian/Italian and South-East Asian/Australian samples ($n = 160$) were assigned to the original population, i.e., the geographic sampling areas. Several population indices were calculated using the ARLEQUIN (ver. 2.000) software (Schneider et al., 2000), testing different population structures, i.e., the geographic population and different population structures suggested by our MJ network analysis.

On intra-population level the haplotype diversity H and nucleotide diversity π, as well as their standard deviation, were determined. Genetic distances between the mitochondrial haplotypes were calculated under the Tamura and Nei model (Tamura and Nei, 1993) after excluding character states with indels. Only loci with less than 5% missing data were used for analysis. The well-known heterogeneity of substitutions across the D-loop region was taken into account and the gamma distribution rate (Yang, 1994) of $\alpha = 0.20$ (± 0.04), estimated by the maximum likelihood approach implemented in TREE-PUZZLE 5.0 (Schmidt et al., 2002), was employed.

Population pairwise F_{ST} values were calculated and significance at the 5% level determined after 110 permutation steps. With ARLEQUIN the distribution of the observed number of differences between all possible pairs of haplotypes (mismatch distribution) was calculated. Usually the distribution reflects the stochastic shape of gene trees, thus is multimodal in samples drawn from population at demographic equilibrium, but is usually unimodal in populations having passed through a recent demographic expansion (Rogers and Harpending, 1992; Slatkin and Hudson, 1991). The goodness-of-fit of the observed data to the simulated model of expansion was tested using the sum of square deviations (SSD) and the raggedness index r (Harpending, 1994), as implemented in ARLEQUIN.

3. Results

3.1. MtDNA variations

The mitochondrial D-loop DNA sequence was determined for 80 water buffaloes, i.e., 19 swamp buffaloes (Carabao) and 61 river buffaloes (24 Murrah, 20 Jafarabadi, and 17 Mediterranean). Abbreviations for these

breeds are Car, Mur, Jaf, and Med, respectively, combined with individual numbers (Table 1). As shown for other mammals (Saccone et al., 1987), the D-loop region of buffaloes is rich in A/T nucleotides. We found that river buffaloes have a slightly higher A/T content (Mediterranean 59.93%, Murrah 59.76%, Jafarabadi 59.86% with an overall standard deviation of 0.2%) than Carabaos ($58.30 \pm 0.16\%$). Nevertheless, these values are smaller compared to those of 42 randomly from GenBank chosen D-loop sequences of cattle ($61.76 \pm 0.16\%$).

The DNA sequence comparison revealed 36 different mitochondrial haplotypes with 128 polymorphic sites. The complete sequence alignment has been deposited in NCBI GenBank (PopSet Accession No. 28631405). An alignment of only the variable positions is shown in Fig. 1. One mitochondrial type was usually shared by animals of the same breed, if it was not unique to one animal. However, there were some exceptions (see also Fig. 3b). The DNA sequence of the haplotype Jaf-02 was shared by individuals of all three investigated river buffalo breeds, i.e., Jafarabadi ($n = 9$), Murrah ($n = 3$), and Mediterranean ($n = 2$). Moreover, this haplotype appears with the highest frequency (14/80). The haplotypes Med-C02, Mur-IZ192, and Mur-310 were found both in Murrah and Mediterranean breeds. On the other hand, the sequence of Mur-481 differs from that of all other river buffaloes, clearly observable in the sequence alignment (Fig. 1) and demonstrated graphically in the network analyses (Figs. 3 and 5). The diversity within the Carabao breed was extremely low, differing just in length polymorphisms and one transition in Car-370 at position 16044 of the alignment (all details about character positions in the alignment are with respect to the bovine sequence, Anderson et al., 1982).

The DNA sequences comparison of the complete alignment revealed 48 $T \leftrightarrow C$ and 37 $A \leftrightarrow G$ transitions, but only nine transversions, demonstrating the strong bias towards transitions. Five positions in the alignment display three-character status (positions 15945, 15965, 16047, 180, and 353 in Fig. 1).

Within the river and swamp buffaloes we observed three regions displaying length polymorphisms, indicated by asterisks below the sequence in Fig. 2. In the 3′-region close to the ETAS1, a poly-G motif harbours 4–6 guanines in the river buffaloes, including Mur-481, and 10–12 guanines in the Carabaos. In the 5′-region two poly-C stretches were detected. The first stretch consists of 7 or 8 and the second one of 7–9 cytosines.

As this is to our knowledge the first study on the complete bubaline D-loop region we generated a D-loop consensus sequence for both river and swamp buffaloes [during the end of the revision process of this manuscript a set of *B. bubalis* mitochondrial DNA sequences was deposited in NCBI GenBank by Sandhu and Kumar (unpublished)]. However, our dataset of swamp buffaloes (Carabao) is most likely not representative, as mentioned above and will be shown in further analyses. Therefore, only the consensus of river buffaloes is given and is compared to the bovine sequence (Fig. 2). In a given position, bases were defined to be the consensus base when they were present in more than 50% of all sequences. The lengths of the bubaline and bovine sequences differ by 17 bp: whereas the bovine D-loop region consists of 909 bp, the bubaline one consists of 926 bp. These differences are due to multiple insertion/deletion (indel) events, mainly in the hypervariable 3′ and 5′ ends of the D-loop region (in Fig. 2 indicated by dashes).

Aligned to the bovine sequence we evaluated characteristics of the D-loop region, e.g., extended termination associated sequences (ETAS) and conserved sequence blocks (CSB) (Saccone et al., 1987; Sbisa et al., 1997). The similarity between cattle and buffalo in the ETAS1 and two regions is 83 and 70%, respectively. Whereas the CSB1 is highly conserved between the species (92%), the bubaline CSB2 is characterized by an insertion of five nucleotides (65% similarity). The conserved central domain (CD) is conserved with a similarity of 89% between the species. Other characteristic elements of the D-loop, i.e., the origin of heavy strand replication (O_H), light and heavy strand promoters (LSP, HSP), are noted in Fig. 2 based on homology comparisons with other species (Anderson et al., 1982; Wood and Phua, 1996). The 5′-end of the bovine sequence consists of a C_{12} stretch, whereas the buffalo sequences do not display such a motif.

3.2. Maximum likelihood analysis of Brazilian/Italian samples

The ML-tree (Fig. 3a), rooted by *Bos taurus*, shows two main branches, i.e., one branch encompasses all individuals of Carabao and Mur-481 and the other one, river buffaloes only. Taking the sequence differences shown in Fig. 1 into account, we expected that Mur-481 (phenotype: river buffalo) would branch together with individuals of the Carabao breed. Hence, the RP of almost 100% demonstrate a high consistency that (a) there is a clear branching into river buffalo and Carabao and (b) the Mur-481 is closely related to the latter. The Carabao individuals itself are less divers and were identified as identical sequences (except Car-370) in the TREE-PUZZLE 5.0 calculation. Therefore, in further analyses we are going to treat this population as two haplotypes.

The group of river buffaloes itself is fairly diverse, however under the ML model many sequences were identical and thus displaying a lack of resolution, resulting in a "star-like" branching (Fig. 3a). Still, the upper branch harbours the groups [[Mur-310, Med-C04, Med-D04, Med-D07], Mur-181] (RP = 90), [Jaf-01, Jaf-25] (RP = 80), [Med-C02, Med-D05] (RP = 55), and

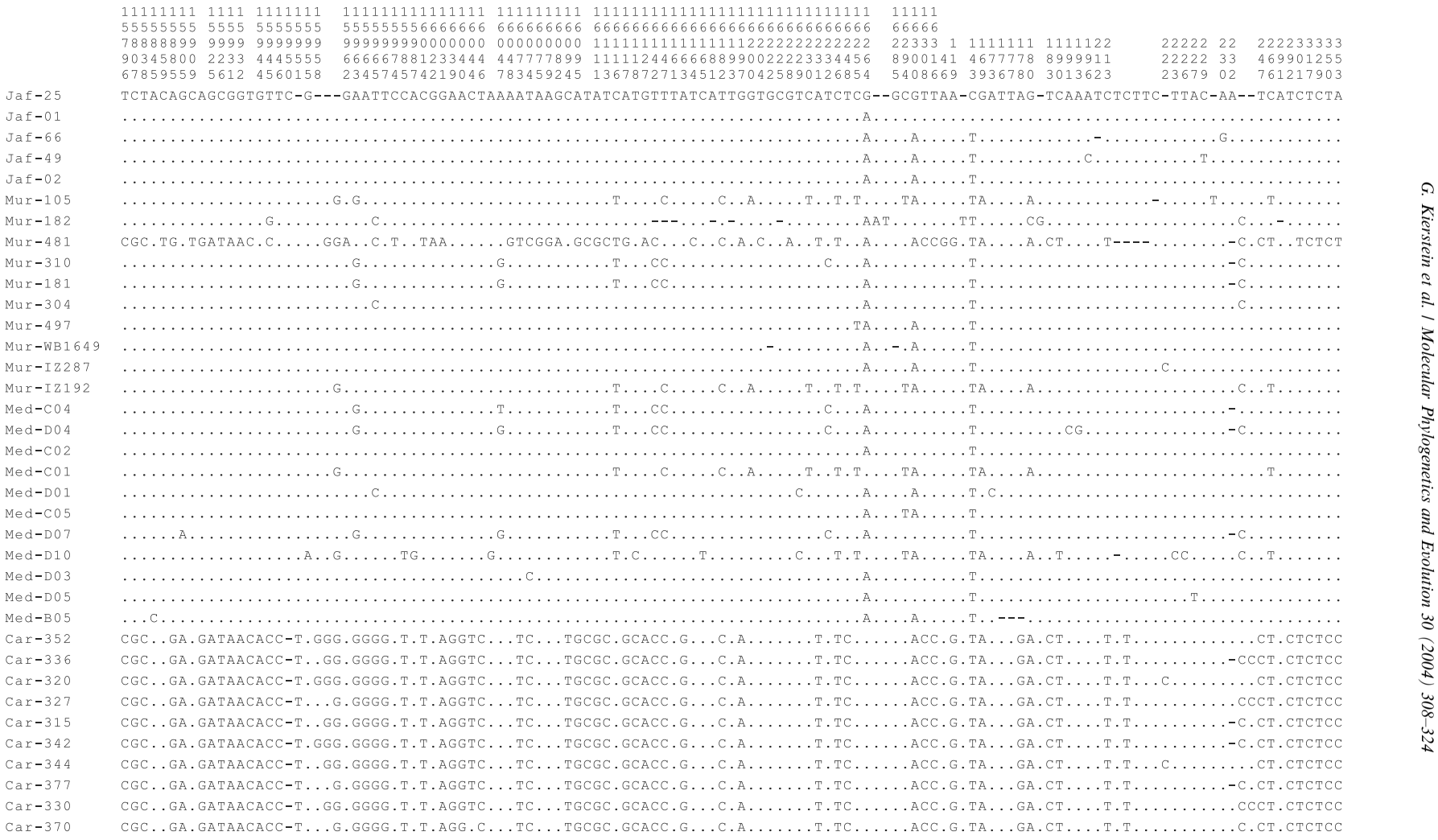

Fig. 1. Mitochondrial D-loop region DNA sequence variations detected in 80 Brazilian and Italian water buffaloes. DNA sequences of the D-loop region of Brazilian river and swamp buffaloes (Jaf, Jafarabadi; Mur, Murrah; Car, Carabao) and Italian river buffaloes (Med, Mediterranean) were determined as described and aligned. DNA sequences are aligned to the sequence of Jaf-25. Differing nucleotides are noted. Identical positions are indicated by dots, deletions by dashes. Nucleotide positions correspond to the bovine sequence (Anderson et al., 1982). The complete sequence alignment has been deposited in NCBI GenBank's PopSet (Accession No. 28631405).

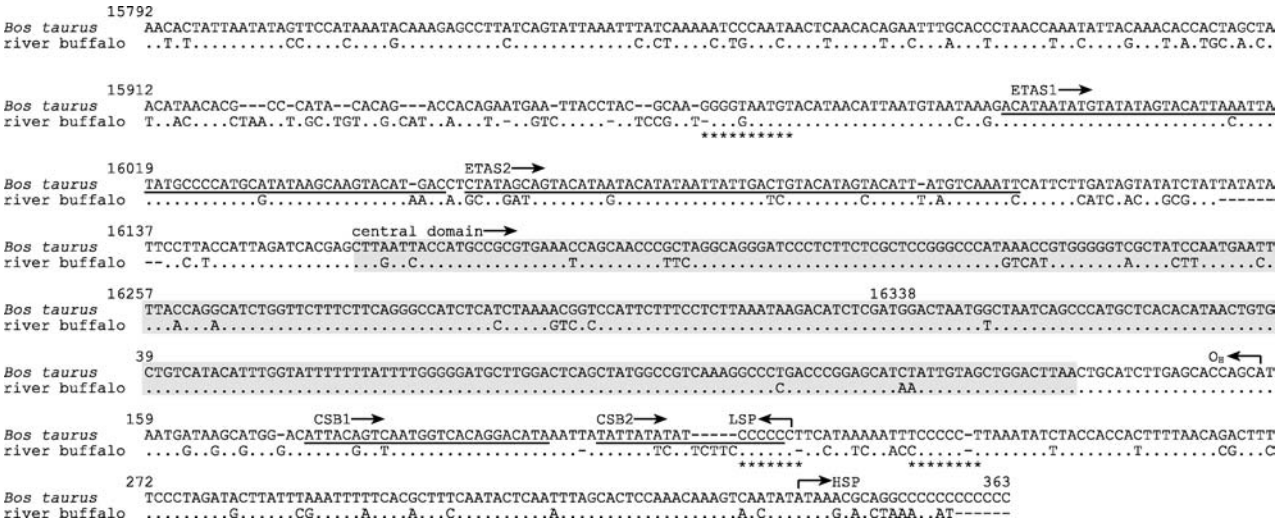

Fig. 2. Comparison of variable sites in the D-loop region of the consensus sequence of river buffaloes aligned to the sequence of *B. taurus*. Structural elements of the D-loop region are underlined (ETAS1, ETAS2, CSB1, CSB2). The conserved central domain is grey shaded. The H-strand origin of replication (O_H), the L-strand promoter (LSP), and the H-strand promoter (HSP) are indicated. Nucleotide positions above the first character in each row refer to the bovine mtDNA sequence (Anderson et al., 1982). Asterisks (*) below the sequence indicate regions of bubaline length polymorphisms.

[[Mur-182, Mur304], Med-D01] (RP = 54). Separated (RP = 71) from these river buffaloes is a group of two Mediterranean and two Murrah haplotypes [[Med-C01, Mur-IZ192, Mur105], Med-D10].

3.3. Network analysis of the Brazilian/Italian samples

For the MJ network analysis of our data we weighed the character states as described in Section 2 and used as a first approach a value of $\varepsilon = 0$ resulting in a network (Figs. 3b and c) with 10 median vectors (*mv*), four within the group of river buffaloes three separating the river buffaloes from the group [Carabao, Mur-481], and three within the Carabao. This network in fact shows similar relationships we found in the ML tree, but here more details can be recognized. The main predictions of the ML-tree are repeated in the network graph: the clear separation of river buffaloes (Fig. 3b) from the Carabao breed (Fig. 3c) and the different "mitochondrial" history of Mur-481 compared to all other river buffaloes analysed here. Beyond that, more detailed information can be seen in the network graph. The above-described group [[Mur-310, Med-C04, Med-D04, Med-D07], Mur-181] displays an relationship with haplotype Mur-310 (found in Murrah and Mediterranean breeds) having mutational connections to four singletons, individuals of Murrah as well as Mediterranean buffaloes. Interesting is the relationship in the magnified part of Fig. 3b. Jaf-02 has direct mutational connections to seven other haplotypes and, like Jaf-02 itself is found in three different breeds, it has connections to all of these three. These kinds of haplotypes, which are related by mutational steps to more than one haplotype and usually displaying a high frequency, are referred to as interior or ancestral haplotypes (Posada and Crandall, 2001).

Directly connected to Jaf-02 is Med-C02, a haplotype found in Mediterranean and Murrah breeds, which itself displays mutational links to singletons, i.e., Med-D05, Med-D03, and Jaf-01 (in line with Jaf-25). Although Jaf-01 demonstrates a high frequency of Jafarabadi ($n = 7$) one has to take into account that these were raised on only one farm in Brazil (Table 1) and thus this haplotype frequency should not be overestimated (the possibility of having sampled sister individuals is very likely). The haplotypes Med-C02 and Jaf-02 are linked to a median vector (*mv3*), from where connections rise (i) to the group of haplotypes around Mur-310, (ii) to Mur-304 and further on to Mur-182, and (iii) to Med-D01. Another group of river buffaloes [Med-C01, Mur-IZ192, Mur-105, and more distant Med-D10] is linked to the Jaf-02 group through a median vector (*mv4*) and Mur-497 and Med-C05, respectively.

The Carabao and Mur-481 haplotypes are clearly separated from the river buffaloes (Fig. 3c) as was suggested already by the ML-tree (Fig. 3a). The higher resolution of the Carabao haplotypes here is due to the fact that gaps were treated as one evolutionary event.

When excluding gaps from the analysis the network basically remains the same with some exceptions, which are shown in Fig. 4 for the river buffaloes only. Not illustrated is the reduction of the Carabao branch to two haplotypes (Car-370 with $n = 1$ and Car-315 with $n = 18$). Within the river buffaloes three median vectors (*mv3*, *mv6*, and *mv7* in Fig. 3b) are resolved leading to a

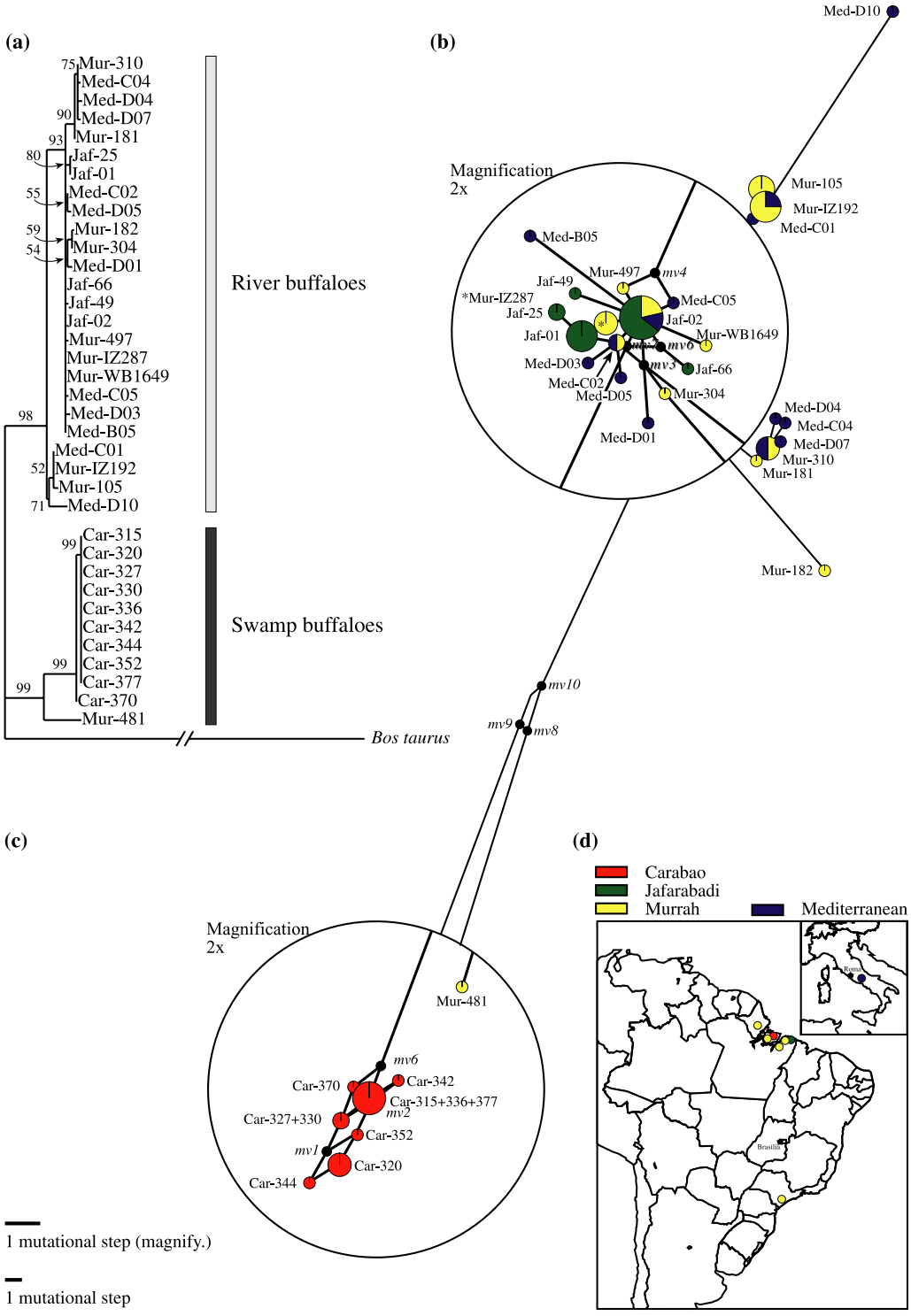

Fig. 3. Phylogenetic relationship among domestic water buffaloes (*Bubalus bubalis*) from four different breeds sampled in Brazil and Italy (d). (a) Maximum likelihood (ML) phylogram reconstructed by TREE-PUZZLE 5.0 from 36 Brazilian and Italian haplotypes of the mitochondrial D-loop region, rooted by *Bos taurus* (Anderson et al., 1982). Quartet puzzling (QP) support after 10,000 puzzling steps are reported on the nodes as reliability percentages (RP). (b) and (c) Median-joining (MJ) network ($\varepsilon = 0$) for the same haplotypes based on the polymorphic sites of the mitochondrial D-loop region. Circled areas are all proportional to the frequency of the haplotypes indicated. If individuals of different breeds shared one haplotype, a pie illustrates the respective proportions. Please note that (b) and (c) are parts of one network with "magnified" parts (big encircles), which were hardly legible in original size, however only the branch lengths were affected, not the areas of haplotype circles. Median vectors (*mv*), produced by the network software, representing missing or not sampled haplotypes, are illustrated by small dots (●).

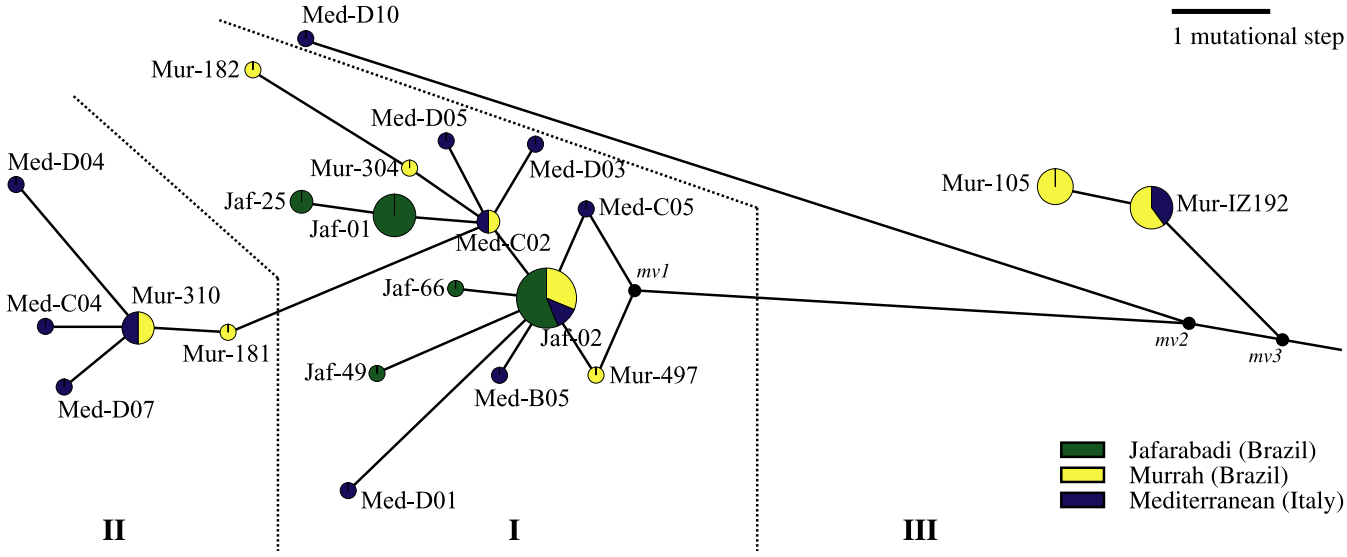

Fig. 4. MJ network as in Fig. 3b, but deletions were excluded from the analysis. We divided the haplotypes into three groups (I, II, and III) for a differentiated calculation of average pairwise distances (Table 3). Group I consists of all three analysed breeds (Jaf, Mur, and Med), whereas the other groups do not inhabit Jafarabadi, but consist mainly of Mediterranean (II) or Murrah (III).

less complex network graph. The group of haplotypes around Mur-310 is now directly linked to Med-C02, like Mur-304 and Mur-182.

When increasing the ε-value to obtain more median vectors (Bandelt et al., 1999), we obtained a slightly more complex network, which provided no additional information (data not shown). Basically, additional median vectors separated Med-D10 and Mur-105, Mur-IZ192 from the other river buffaloes by three cuboids. A prism was generated between Mur-310, Mur-181, and Med-C04, thereby separating the group from Med-C02 by additional median vectors.

Based on the network graph of Figs. 3b and 4 we divided the river buffaloes into three groups (I, II, and III, see Fig. 4); a central group I with a prominent haplotype found in all three river buffalo breeds and connections to singletons of all breeds, group II descending from group I with a longer mutational link and group III, which is only remotely related to these two groups.

The average pairwise sequence difference was calculated between river and swamp buffaloes (Table 3) using the Tamura and Nei model with gamma distribution (Tamura and Nei, 1993; Yang, 1994). Based on our findings as described above we subdivided the river buffaloes into three groups showing low diversity within the groups I and II (around 0.2%) and higher diversity within group III (almost 1%). Between groups the diversity ranges between 0.9% and almost 1.5%. The average diversity within the river buffaloes is around 0.8%. The divergence between river and swamp buffaloes was calculated to be 13%, however one has to

Table 3
Average pairwise sequence divergence (±SD) between river and swamp buffaloes (Brazilian/Italian samples)

TrN + Γ (uncorrected *p*)		River buffalo			Swamp buffalo
		I	II	III	
River buffalo	I	0.0022 ± 0.0024 (0.0021 ± 0.0023)			
	II	0.0145 ± 0.0024 (0.0139 ± 0.0022)	0.0025 ± 0.0020 (0.0024 ± 0.0019)		
	III	0.0091 ± 0.0071 (0.0084 ± 0.0065)	0.0136 ± 0.0085 (0.0128 ± 0.0074)	0.0096 ± 0.0085 (0.0091 ± 0.0079)	
			0.0078 ± 0.0070 (0.0075 ± 0.0065)		
Swamp buffalo			0.1317 ± 0.0034 (0.0691 ± 0.0024)		0.0173 ± 0.0358 (0.0073 ± 0.0153)

A detailed distance matrix for three groups (I, II, and III) of river buffaloes is given; please refer to Fig. 4. Values are corrected to the Tamura and Nei model plus gamma distribution (TrN + Γ). Uncorrected *p* values are given in parenthesis.

take into account the small number of sampled swamp haplotypes.

3.4. Maximum likelihood analysis of Brazilian/Italian and South-East Asian/Australian samples

As shown and mentioned above, our dataset of Carabaos is not very diverse. Yet we wanted to prove how the Mur-481 haplotype is related to other swamp buffaloes. Therefore, we compared our data with published D-loop region DNA sequences of water buffaloes, raised in Sri Lanka, Thailand, Malaysia, Indonesia, Philippines, and Australia (Lau et al., 1998). Lau and colleagues analyzed a 158 bp fragment of the hypervariable region of the mitochondrial D-loop region between position 16089 and 16241, harbouring the end of ETAS2 and the beginning of the conserved central domain. For a complete alignment we then truncated our sequences. As a consequence the Brazilian/Italian 36 haplotypes were reduced to 10, i.e., nine river buffaloes (including Mur-481) and one haplotype of Carabao. Interestingly, the truncated sequence of haplotype Mur-105 is identical to the South-East Asian/Australian haplotype Aust 1, therefore referred to as "Mur-105 = Aust 1." When conducting the ML analysis the tree could not be completely resolved and there was no separation between river and swamp buffaloes as we have observed in the Brazilian/Italian haplotypes. Out of 41 haplotypes 24 are connected to the same internal node, whereas the other haplotypes are grouped in two to four (with reliability percentages between 62 and 97), which then are linked to the one internal node (Fig. 5a).

Basically the same tree with low resolution was obtained when conducting the ML analysis with the South-East Asian/Australian samples alone (data not shown).

3.5. Median-joining network analysis of the Brazilian/Italian and South-East Asian/Australian samples

The analysis of the pooled samples is therefore based on 160 individual sequences leading to a total of 41 haplotypes.

In contrast to the ML-tree, the MJ network analysis of the pooled South-East Asian/Australian haplotypes (grey coded circles in Figs. 5b and c, Lau et al., 1998) and our data (colour coded circles in Figs. 5b and c) revealed a detailed network. Please note that both figures are connected through median vector *mv6* to one single network. Due to the different kind of linkage among haplotypes, thus the different appearance of both "sub-nets," we refer to them as "central group" (Fig. 5c) and "eccentric group" (Fig. 5b). The eccentric group only consists of swamp buffaloes (compare with Table 4 in Lau et al., 1998) including the Carabao and Mur-481 haplotypes from Brazil. To avoid a misleading interpretation of haplotype frequencies we treated the 19 Carabao samples as having a frequency of $n = 2$ (as mentioned above). The eccentric group displays Phil 2 as the most prominent haplotype with the highest frequency here, found in different populations and with three direct mutational links (Sula 2, Sara 1, and Phil 7). Further, it is linked to other haplotypes (e.g., Carabao, Thai 2, etc.) through median vectors.

The central group consists mainly of phenotypic river buffaloes, i.e., the Brazilian/Italian samples and the three lightest grey coded South-East Asian/Australian samples, i.e., Sri Lankan Lankan buffalo (SriL), Sri Lankan Murrah (SriL M), and Malaysian Murrah (Mal M) as indicated by the small map in Fig. 5c. However, we also find swamp buffaloes like Thai 1 close related to river buffaloes (SriL M1, Mur-181) and even haplotypes (Aust 1 and Aust 2), which are found in both river and swamp buffaloes. This is in agreement with findings of Lau and colleagues (1998).

Compared to the MJ network of Brazilian/Italian river buffaloes (Figs. 3b and 4) more median vectors have been added by the program, resulting in one "prism" and one "domino" (two four-cycles sharing a link), with the latter sharing a link with an additional four-cycle.

Several interior haplotypes are found in this network. In the "prism" Phil 3 ($n = 8$), which is found in four populations and has mutational links to two singletons from other populations. All three haplotypes rank among the swamp buffaloes. From Phil 3 two ways, with three or two mutational steps, lead to "Mur-105 = Aust 1": (i) through median vector *mv2* and a Murrah (Mal M2) or (ii) alternatively through *mv8*, which itself has a connection to a Lankan buffalo. Barker and colleagues have demonstrated the close relationship between Lankan and Murrah buffaloes (Barker et al., 1997b). The haplotype "Mur-105 = Aust 1" is found in Brazilian/Italian Murrah and Mediterranean breeds as well as in Lankan ($n = 5$), Sri Lankan Murrah ($n = 1$), and swamp buffaloes ($n = 6$).

From "Mur-105 = Aust 1" through *mv4*, *mv3* the median vector *mv1* is reached. Interestingly, this median vector resembles an interior haplotype (except it has by nature a frequency of $n = 1$), as it has multiple connections to singletons (Mal M3, Sara 2, SriL M2, and Aust 2) and further links to haplotypes (SriL M1, SriL M3), which give rise to others. It is noteworthy that *mv1* links to both river and swamp buffalo haplotypes (note also: Aust 2 ($n = 14$) is found in river and swamp buffaloes).

The haplotype SriL M3 links up to Jaf-02, which exhibits short and longer mutational connections to Murrah, Mediterranean, and one Lankan buffalo, thus this is the only pure river buffalo branch of the network. Additionally, SriL M3 connects also to a swamp buffalo (Thai 1), which is linked to a further swamp buffalo (Phil

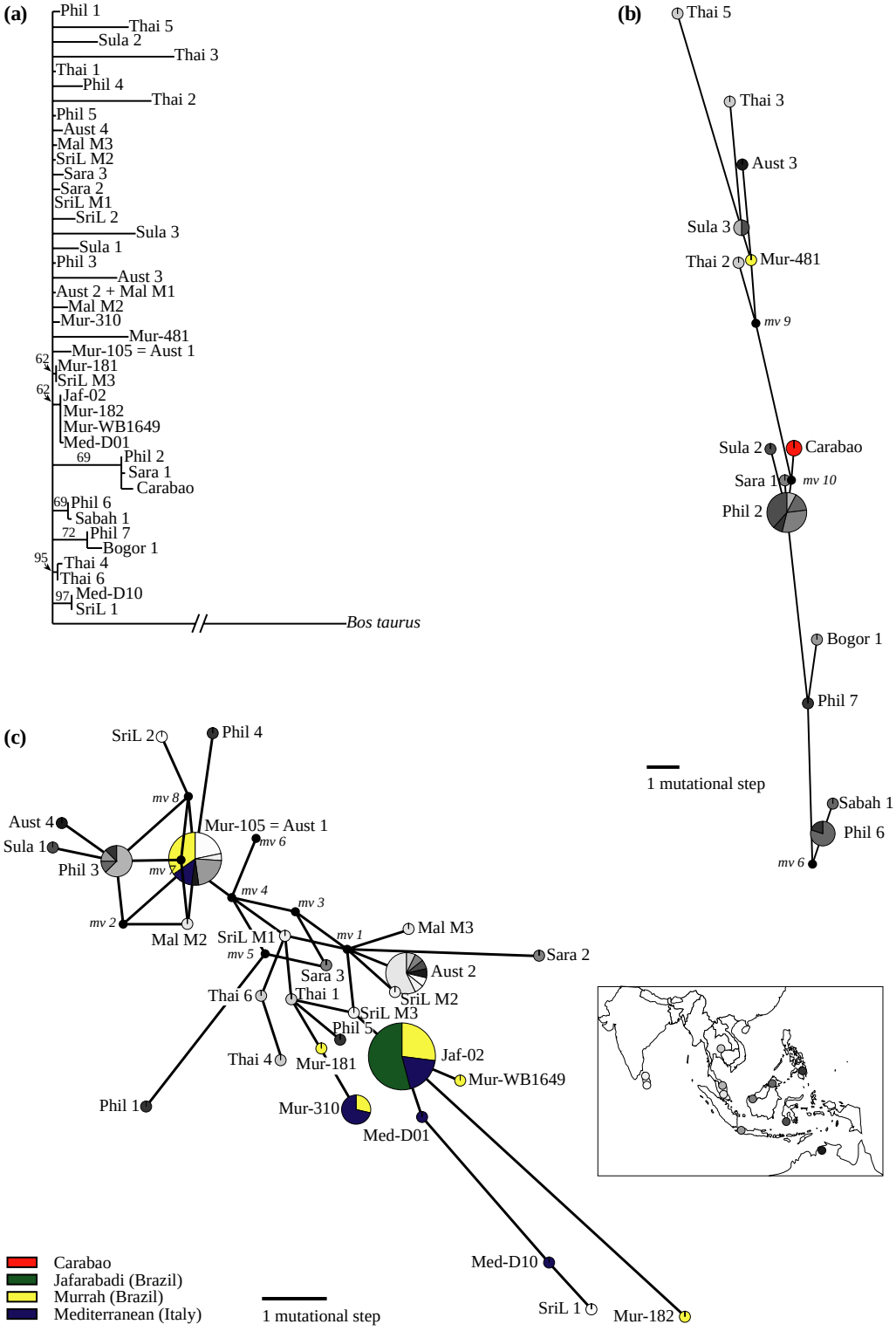

Fig. 5. Phylogenetic relationship among domestic water buffaloes from Southeast Asia (Lau et al., 1998) combined with our Brazilian and Italian data. (a) ML tree reconstructed by TREE-PUZZLE 5.0 based on a D-loop fragment of 141 bp. Quartet puzzling (QP) support after 10,000 puzzling steps are reported on the nodes. (b) and (c) MJ network ($\varepsilon = 0$) for the same set of domestic water buffaloes; Brazilian/Italian populations are colour coded. Circled areas are all proportional to the frequency of the haplotypes indicated, except for "Carabao" which has been corrected to two haplotypes due to biased sampling on one farm (for details see main text body). If individuals of different breeds shared one haplotype, a pie illustrated the respective proportions. Median vectors (*mv*) are indicated by small dots (●). *Note*. The network had been separated at median vector *mv6* into the "eccentric group" (b) and the "central group" (c). The network in (c) has been enlarged (2×) for better clarity of the connections, however only the branch lengths were affected, not the circled areas. Framed is a map where the Southeast Asian/Australian buffaloes were sampled (for details see Barker et al., 1997b).

5) and Brazilian/Italian river buffaloes (Mur-181, Mur-310).

When increasing ε with all character states weighed equally a complex network with so-called large cells was obtained (not shown). However, when increasing ε with character weighed as described in Section 2 the algorithm runs into a continuous cycle.

Summarizing the network analysis results of the pooled data we found one "central group" of haplotypes found in both river and swamp buffaloes. One "eccentric group" consists of haplotypes only found in swamp buffaloes with the exception of Mur-481. However, the relationship among the eccentric group is not star-like as one would expect for a domestication event followed by an expansion. In contrast the central group shows interior haplotypes with star-like connections. Based on this we hypothesize a single domestication of water buffaloes leading to the haplotypes and relationships found in the central group of the MJ network (Fig. 5c). The haplotypes of the eccentric group descent from probably wild water buffaloes crossbred with domestic flocks.

3.6. Population demography

As a further approach different population indices were addressed. To tackle the problem of one or two domestication events of river and swamp buffaloes we compared different groupings of populations and the effect of omitting certain haplotypes, e.g., the "eccentric group" of swamp buffaloes, from the analysis.

The genetic diversity in the different populations is given in Table 4. When excluding the "eccentric group" the nucleotide diversity π in the South-East Asian/Australian populations range between 0.5% in the Malaysian Murrah and around 7% in the Sarawak population. The Brazilian Murrah and the Italian Mediterranean show similar diversity (1.5 and 1.7%), whereas the Brazilian Jafarabadi breed shows virtually no diversity regarding the analysed fragment of 150 bp. Gene diversity H among the South-East Asian/Australian populations range between 0.3 and 1.0 and among the Brazilian/Italian populations between 0.7 and almost 1.0. These results are of course biased by the number of samples in a population (n) and the number of haplotypes (k) found in each population. These values should therefore be interpreted with caution. Including the haplotypes of the "eccentric group" leads in general to higher values of π and H, except for the Sulawesi population, where the ratio of n/k decreased (Table 3).

The population pairwise F_{ST} values (Table 5) are based on those mitochondrial haplotypes, which are found in the "central group." Note this is different from the approach of Lau et al. (1998). The Italian Mediterranean population is significant different from all other populations except the Brazilian Murrah. The Brazilian Jafarabadi is different from all other populations. However, one has to take into account that this population is described by only one haplotype Jaf-02. Within the South-East Asian/Australian populations the Australian population shows no difference to any other population of South-East Asia and is only different to the Italian buffaloes (and the Brazilian Jafarabadi). The Lankan buffalo from Sri Lanka is different from the Sri Lankan and Malaysian (and Brazilian) Murrah, and from the Surin and Sarawak population, but not from the Philippine, Sulawesi, and Bogor swamp buffalo populations. These are different findings compared to

Table 4
Genetic diversity in 14 populations from South-East Asia, Australia, Italy, and Brazil

Populations	Excluding "eccentric group"				Including "eccentric group"			
	n	k	π(SD)	H(SD)	n	k	π(SD)	H(SD)
SriL-Lan	8	4	0.017 (0.012)	0.643 (0.184)	—	—	—	—
SriL-Mur	5	5	0.017 (0.013)	1.000 (0.127)	—	—	—	—
Mal-Mur	10	4	0.006 (0.005)	0.533 (0.180)	—	—	—	—
Surin-S	4	4	0.016 (0.013)	1.000 (0.177)	7	7	0.097 (0.057)	1.000 (0.076)
Trengg-S	5	1	0.0 (0.0)	0.0 (0.0)	7	3	0.074 (0.044)	0.524 (0.209)
Bogor-S	6	2	0.002 (0.003)	0.333 (0.215)	7	3	0.016 (0.012)	0.524 (0.209)
Sara-S	3	3	0.070 (0.056)	1.000 (0.272)	8	5	0.086 (0.049)	0.786 (0.151)
Sabah-S	1	1	0.0 (0.0)	0.0 (0.0)	8	4	0.044 (0.027)	0.750 (0.139)
Phil-S	4	4	0.031 (0.023)	1.000 (0.177)	7	7	0.025 (0.017)	1.000 (0.076)
Sula-S	2	2	0.064 (0.067)	1.000 (0.500)	9	5	0.062 (0.036)	0.722 (0.159)
Aust-S	3	3	0.014 (0.014)	1.000 (0.272)	4	4	0.079 (0.055)	1.000 (0.177)
Italy-Med	17	14	0.017 (0.011)	0.978 (0.027)	—	—	—	—
Braz-Jaf	20	5	0.0 (0.0)	0.695 (0.070)	—	—	—	—
Braz-Mur	24	11	0.015 (0.010)	0.909 (0.034)	—	—	—	—
Total	112	46	0.021 (0.012)	0.943 (0.011)	112	46	0.021 (0.012)	0.943 (0.011)

Note. n, sample size; k, number of haplotypes; π = nucleotide diversity assuming the Tamura and Nei model (Tamura and Nei, 1993) with gamma distribution $\alpha = 0.24$ (Hasegawa and Kishino, 1989); H, gene diversity (Nei, 1987). Abbreviation used for populations/countries and breeds: SriL (Sri Lanka), Mal (Malaysia), Trengg (Trengganu), Sara (Sarawak), Phil (Philippines), Sula (Sulawesi), Aust (Australia), Braz (Brazil), and Lan (Lankan), Mur (Murrah), Med (Mediterranean), Jaf (Jafarabadi), and S (swamp buffalo).

Table 5
Population pairwise F_{ST} values (below the diagonal) and average pairwise differences within population (underlined)

	SriL-Lan	SriL-Mur	Mal-Mur	Surin-S	Bogor-S	Sarawak-S	Phil-S	Sulawesi-S	Australia-S	Italy-Med	Brazil-Jaf	Brazil-Mur
SriL-Lan	2.46											
SriL-Mur	**0.22**	2.40										
Mal-Mur	**0.51**	**0.18**	0.80									
Surin-S	**0.36**	−0.06	**0.34**	2.17								
Bogor-S	−0.03	**0.53**	**0.78**	**0.68**	0.33							
Sarawak-S	**0.35**	0.07	**0.35**	0.04	**0.49**	10.00						
Philippines-S	0.16	0.06	0.44	0.10	0.34	−0.02	5.00					
Sulawesi-S	0.27	0.16	0.49	0.20	0.45	−0.05	−0.17	9.00				
Australia-S	−0.18	−0.05	0.39	0.16	0.13	0.04	−0.06	−0.32	2.33			
Italy-Med	**0.32**	**0.17**	**0.49**	**0.26**	**0.49**	**0.41**	**0.29**	**0.45**	**0.25**	2.35		
Brazil-Jaf	**0.81**	**0.79**	**0.92**	**0.87**	**0.98**	**0.79**	**0.79**	**0.90**	**0.92**	**0.28**	0.00	
Brazil-Mur	**0.24**	0.18	**0.48**	**0.30**	**0.38**	**0.46**	**0.28**	**0.46**	0.16	0.02	**0.33**	2.11

Note. Significant F_{ST} values at the 5% level are in bold. Abbreviation used for countries and breeds: SriL (Sri Lanka), Mal (Malaysia), Phil (Philippines), and Lan (Lankan), Mur (Murrah), Med (Mediterranean), Jaf (Jafarabadi), and S (swamp buffalo).

those of Lau and colleagues who found significant differences between Lankan buffaloes and the Philippine and Sulawesi populations (1998). The Sri Lankan Murrah is different from the Malaysian one, and from the Bogor population, but there is no significant difference to all the other population. The Malaysian Murrah is different from both Sri Lankan populations and those of Surin, Bogor, and Sarawak. The Philippine and Sulawesi populations are not significant different from any other population from South-East Asia, not even from the river buffaloes. The Surin population from the South-East Asian mainland differs only from the Lankan and Malaysian Murrah and the swamp buffaloes of Bogor. To summarize, among the swamp buffalo populations only two F_{ST} values demonstrate significant population difference (Bogor–Surin, Bogor–Sarawak). However, among the three river buffalo populations (SriL, SriL M, and Mal M) all pairwise F_{ST} values are significant different. Comparing the river with the swamp buffalo population, two-third of possible pairs show no significant difference.

Additionally a mismatch distribution analysis (Fig. 6) was employed to give further support for a single domestication event. The mismatch distribution among the pooled water buffaloes (Fig. 6a) is multimodal with two main peak areas enclosing one small peak, plus one peak at zero mismatches, and one small peak at high number of sequence mismatches. We removed the Brazilian/Italian haplotypes to have a closer look on the South-East Asian population. The mismatch distribution among the South-East Asian/Australian Swamp buffalo populations as (defined in Lau et al., 1998) is multimodal with two medium sized peaks and several smaller peaks (Fig. 6b), a distribution one would not expect for expanding populations (goodness-of-fit: $SSD = 32.5 \times 10^{-3}$; $r = 0.04$). However, when grouping the haplotypes referring to the network diagram (Figs. 5b and c) into the central and the eccentric groups, the mismatch distribution alters. The central group (Fig. 6c) is bimodal with one prominent and one minor peak ($SSD = 8.32 \times 10^{-3}$; $r = 0.03$). This minor peak is reduced clearly when removing the Lankan haplotypes from this group (Fig. 6d; $SSD = 9.05 \times 10^{-3}$; $r = 0.03$). In contrast, the eccentric group remains multimodal (Fig. 6e; $SSD = 56.01 \times 10^{-3}$; $r = 0.14$), reflecting the long distances between haplotypes in this part of the network, as described above. The goodness-of-fit between the observed mismatch distribution and the simulated expansion model clearly favours the central group having a single origin, rather than the swamp buffaloes.

4. Discussion

4.1. D-loop variation in Brazilian/Italian populations

Comparing all Carabao D-loop sequences we observed four polymorphic sites, three of them due to length polymorphisms. We found one transition at position 16044 in Car-370. The low degree of variability within the D-loop region of the Carabao set is most likely due to the fact that all animals were raised on one farm and probably belong to no more than one or two maternal lineages. This is supported by the fact that when excluding the sites with length polymorphisms, only 2 haplotypes (distinguishable by a C or T at position 16044) are found. The observed G-stretch therefore might be a unique characteristic of this lineage, as the putative swamp buffalo Mur-481 does not show this poly-guanine motif. The "G-stretch" lies outside the sequence included in the analysis of Lau and colleagues, thus it was not possible to compare it with these populations. The deviated lineage of Mur-481 from the river buffaloes, however, is supported by the comparison with swamp buffaloes using the data of Lau et al. (1998). The Brazilian river buffaloes showed a mean diversity. Although the Murrah breed seems to be more diverse, one has to take into consideration that they were sampled on five different farms. In contrast, the Mediterranean samples showed the highest variability. Samples of three different sites (indicated by the capital letter) revealed 11

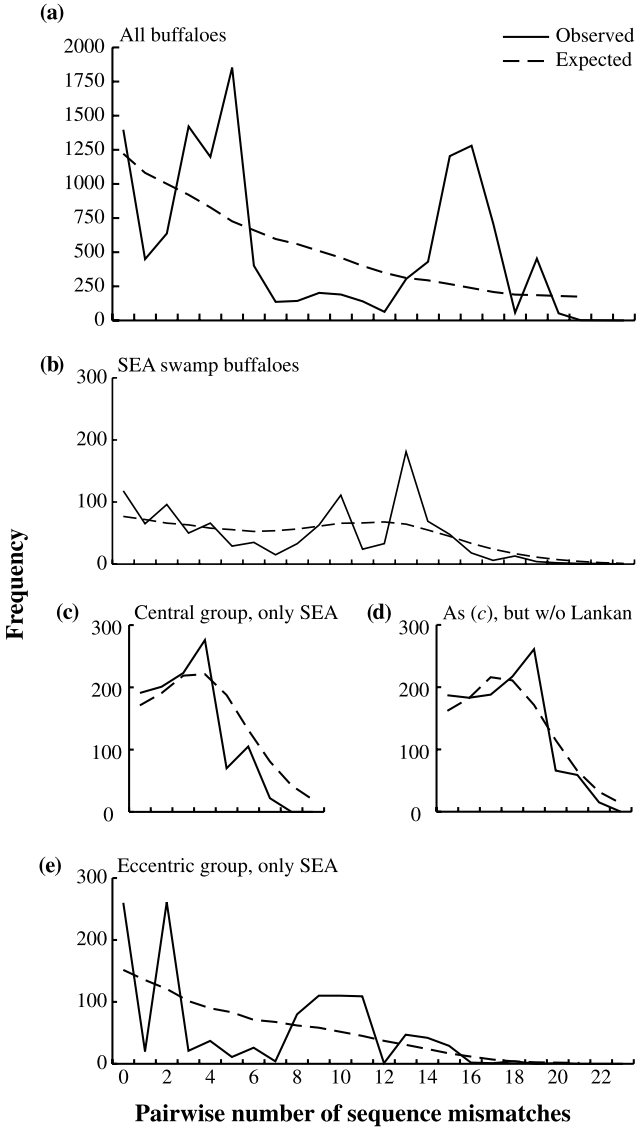

Fig. 6. Mismatch distribution between mitochondrial D-loop haplotypes of water buffalo. This figure is based on the molecular distance (Tamura and Nei model, Tamura and Nei, 1993) between all possible pairs of D-loop haplotypes (150 bp) after pooling following populations: (a) all populations, with a total of 160 sequences; (b) all swamp buffalo populations from South-East Asia/Australia; (c) all haplotypes found in the "central group" of Fig. 5c, excluding the Brazilian/Italian populations; (d) same as (c), but excluding the Lankan population; (e) all haplotypes belonging to the "eccentrics group" as described in Fig. 5b.

different haplotypes out of 17 samples. This result can be expected, as the Brazilian population is founded on relatively small numbers of importations. However, when adjusting our samples to the sequences obtained by Lau et al. (1998), we got only eight haplotypes. Bottleneck and founder effects can explain these extreme differences in diversity. Whereas the Brazilian population is the most recent one, the Asian population is supposed to be the oldest one, thus showing more diversity. It has to be noted that the Brazilian buffalo population derived from an unknown diverse ancestor population, mainly imported from India and Italy starting in the end of the 19th century. Therefore, it is expected to find a lower diversity in the Brazilian population than in the Asian. Mediterranean buffaloes were probably the first buffaloes introduced to Brazil already in 1895 followed by more importations (Fonseca, 1977). In 1962 the first Murrah animals from India were introduced to Brazil and were widely dispersed (Cockrill and Mahadevan, 1974; Nascimento et al., 1979). Thus there might be a high probability of cross-breeding between these two breeds in Brazil, especially under the light that the river buffaloes of India and Italy, which are dark-skinned, are referred to as Preto (black) in Brazil (Cockrill and Mahadevan, 1974).

In our first survey of the Brazilian population we included only a few farms. Further collection of samples of different buffalo populations, including all breeds, is required to increase the knowledge about Brazilian water buffalo diversity.

4.2. Phylogeny inferred from MJ networks and population genetic indices

Using two different methods for showing the intraspecies relationship (bifurcate tree and MJ network) demonstrates the problem in inferring phylogenies on population level using trees. By nature the genetic loci sampled from individuals of one species are different from those sampled from different species. Different species are defined by their reproductive isolation over a long time, thus leading to fixation of alleles and, in due course, to divergent gene pools. By contrast, gene relationship within a species is not hierarchical, as sexual reproduction is still distributing alleles "horizontally." Therefore, the traditional methods developed for inferring phylogenies above species or even genus level do not properly take into account that at population level few phenomena abuse some of the implemented assumptions (for review on intraspecific gene evolution, see Posada and Crandall, 2001).

The evidence for Jaf-02, Phil 3, and Aust 1 (= Mur-310) and perhaps even Aust 2 being ancestral haplotypes can be explained by predictions from the coalescent theory (Hudson and Kaplan, 1995; Watterson and Guess, 1977). The first three haplotypes are found in at least three different populations with a broad geographic distribution, high frequency and links to singeltons. It is intuitively reasonable that there is a direct relationship between the age of one genetic haplotype (allele) and the frequency with which it appears in a population. As a consequence, it is more likely that new mutations arise from broadly distributed haplotypes than from rare ones. Those old haplotypes become so-called interior haplotypes, see Figs. 3–5, having a high frequency and having many mutational connections to singletons.

Those old haplotypes, especially in domestic animals, are getting distributed over large geographical areas. Domestic animals are not strictly restricted to their native habitat; in contrast they have to migrate with human. The broadest geographic distribution is observed in haplotype Aust 2, which is found mainly in Malaysia, but also in Brazil, Italy, Sri Lanka, Thailand, Indonesia, and Australia. Moreover, this haplotype is found in both river and swamp buffaloes. Interestingly, despite its numerical and geographical predominance suggesting the ancestral status, it is found at a terminal position of the network and has no connections to singletons. This is unexpected taking into account the coalescent theory as described above. However, Cockrill (1974) describes in the chapter "the buffaloes of Indonesia" ritual sacrifices of large numbers of buffaloes, leading to "near-bankruptcy" of a whole village and a stock of buffaloes may need years for recovery. These extreme bottleneck and founder effects could explain the "dead-end" position of Aust 2, at least for the Malaysian/Indonesian populations.

The most important finding of this study is the result of the network analysis (Figs. 5b and c) clearly demonstrating no evidence for two domestication events as it has been shown in other species like sheep, pig and cattle (Giuffra et al., 2000; Hiendleder et al., 1998; Troy et al., 2001), with a clear separation in two clades. We, in contrast, found one central group (or clade) with the majority of haplotypes, including both the river and swamp buffaloes. On the other hand we found an outstretching, eccentric branch with only swamp buffalo haplotypes placed there like pearls on a string. Note, these haplotypes of the eccentric branch are only found in the South-East Asia, not in the Indian or Italian population. We would like to stress the gentle transition between the central clade and the eccentric branch, giving no evidence for a second clade. Moreover, the haplotypes along the outstretching branch are on average separated by longer mutational connections than the haplotypes within the central group and displaying no real star-like phylogeny as would be expected for expanding population (Harpending, 1994; Richards et al., 1998). These findings are supported by the mismatch distribution analysis, which shows an almost unimodal distribution for the central group, supporting the recent expansion model, but a multimodal distribution for the eccentric branch. The haplotypes of the eccentric branch can be explained by occasional introgression of wild water buffaloes into the domestic flocks. Further, the F_{ST} values indicate, with some exceptions, no clear separation between the Lankan/Murrah populations and the swamp buffalo population.

Summarizing these finding we found molecular evidence for only one domestication event of water buffaloes. This is additionally supported by Lau and colleagues (1998) who hypothesized one domestication based on their analyses, among others the recent divergence of river and swamp buffaloes in accordance with the presumed time of domestication.

Yet, based on our results we hypothesize a slightly different scenario of water buffalo domestication than given by Lau and colleagues (1998). We propose the water buffalo domestication on the Indian subcontinent, supported by representations of tame (probably domesticated) buffaloes found on seals in the Indus valley and Mesopotamia (Cockrill, 1974). These resembled the today swamp buffaloes and the wild water buffalo. During migration to Indochina and South-East Asia occasional cross-breeding with wild buffaloes could have lead to those haplotypes now found in the eccentric group, intermixing with the haplotypes of the founder population of domestic water buffaloes. A second alternative explaining the origin of the haplotypes of the eccentric branch is based on an earlier (probably independent) domestication of buffaloes in China (Chen and Li, 1989) with each province having its own buffalo strain (Chunxi and Zhongquan, 2001). Chinese colonizer (from different provinces) most likely came with their buffaloes when spreading the rice cultivation to South-East Asia, explaining the longer mutational connections between these haplotypes and the low frequency. Yet to date no genetic assessment of the Chinese local swamp buffalo strains were conducted, which could clarify the relationship between those and the South-East Asian swamp buffaloes.

In India, and later on further west, the today river buffalo due to selection breeding achieving improved breeds with prominent horn forms and high milk production replaced the ancient swamp buffalo phenotype. In contrast, the swamp buffalo of South-East Asian has not been divided in different breeds and still it is common practice in the Indochina region to cross-breed domestic and wild water buffalo to maintain the similarity (Cockrill, 1974). This evolution of the river buffaloes from an ancestral swamp-like buffalo has already be suggested by Lau and colleagues (1998) further supported by the fact that the phenotypic swamp-like Lankan buffalo is genetically a river buffalo (Barker et al., 1997a,b). However, Lau et al. (1998) suggested the domestication origin in the area of the South-East Asian mainland with a subsequent spread to the north (China) and west (India). If this were true one would expect the highest diversity in the Asian mainland with decreasing diversity moving away from the centre of origin (Loftus et al., 1999). This has been understood to be a sign of human exportation of selected domestic lineages, with subsequent minimal effects from processes like mutation and drift over the brief time period involved. Our analysis revealed only for the Sulawesi, Sarawak, and Philippine population higher nucleotide diversity than the Indian population (Table 4, excluding eccentric group). The Philippine swamp buffaloes have been

introduced probably by Malay immigrants already more than 2200 years ago, later on by Chinese colonizers. Since the 1920s subsequent importations of Indian Murrahs aimed to upgrade the milk and meat production of the native swamp buffaloes (Coronel, 1974). This continuous input of new genetic material can count for the high diversity of the Philippine population. All other South-East Asian/Australian population have lower diversity than the two Indian populations. Although the sample sizes a quite small and therefore the values are biased they support the Indian origin.

In conclusion our results support the following scenario of water buffalo domestication and migration. Domestication of water buffaloes occurred in the Indian subcontinent 5000 years ago. On the South-East Asian mainland these population interbred with wild buffaloes and/or domestic animals from China. There might have been two migration routs onto the islands of Oceania. One from the mainland south to Thailand and Malaysia, another one going up to China and spreading with rice cultivation through Taiwan and the Philippines to Indonesia.

Acknowledgments

The authors thank EMBRAPA/PRODETAB for financial support of sampling the Brazilian water buffalo breeds. We also thank three anonymous reviewers for their helpful comments on the manuscript, and O. Hanotte and D. Mburu for helpful suggestions and explanations, especially on the population analyses, during the revision process. L. Sena is thanked for the investigation on the sparse information on early history of Brazilian water buffaloes. A. Deppe is thanked for expert technical assistance. This work was partly supported by a grant of the DAAD (Deutscher Akademischer Austauschdienst) to B. Brenig and G. Kierstein.

References

Amano, T., Miyakoshi, Y., Takada, T., Kikkawa, Y., Suzuki, H., 1994. Genetic variants of ribosomal DNA and mitochondrial DNA between swamp and river buffaloes. Anim. Genet. 25 (Suppl. 1), 29–36.

Anderson, S., de Bruijn, M.H., Coulson, A.R., Eperon, I.C., Sanger, F., Young, I.G., 1982. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J. Mol. Biol. 156, 683–717.

Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48.

Barker, J.S., Moore, S.S., Hetzel, D.J., Evans, D., Tan, S.G., Byrne, K., 1997a. Genetic diversity of Asian water buffalo (*Bubalus bubalis*): microsatellite variation and a comparison with protein-coding loci. Anim. Genet. 28, 103–115.

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

Boyce, W.M., Ramey II, R.R., Rodwell, T.C., Rubin, E.S., Singer, R.S., 1999. Population subdivision among desert bighorn sheep (*Ovis canadensis*) ewes revealed by mitochondrial DNA analysis. Mol. Ecol. 8, 99–106.

Chen, Y.S., Li, X.H., 1989. New evidence of the origin and domestication of the Chinese swamp buffalo (*Bubalus bubalis*). Buffalo J. 1, 51–55.

Chunxi, Z., Zhongquan, L., 2001. Buffalo Genetic Resources in China. In: Buffalo Newsletter, pp. 1–7.

Cockrill, W.R., 1974. The Husbandry and Health of the Domestic Buffalo. FAO, Rome.

Cockrill, W.R., 1981. The water buffalo: a review. Br. Vet. J. 137, 8–16.

Cockrill, W.R., Mahadevan, P., 1974. The buffaloes of Latin America. In: Cockrill, W.R. (Ed.), The Husbandry and Health of the Domestic Buffalo. FAO, Rome, pp. 676–707.

Coronel, A.B., 1974. The buffaloes of the Philippines. In: Cockrill, W.R. (Ed.), The Husbandry and Health of the Domestic Buffalo. FAO, Rome, pp. 616–622.

Cymbron, T., Loftus, R.T., Malheiro, M.I., Bradley, D.G., 1999. Mitochondrial sequence variation suggests an African influence in Portuguese cattle. Proc. R. Soc. Lond. B 266, 597–603.

Douzery, E., Randi, E., 1997. The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. Mol. Biol. Evol. 14, 1154–1166.

FAO, UNEP, 2000. World Watch List for Domestic Animal Diversity, 3rd ed. Information Division, FAO, Rome.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Fischer, H., Ulbrich, F., 1968. Chromosmes of the Murrah buffalo and its crossbreeds with the Asiatic swamp buffalo (*Bubalus bubalis*). Z. Tierzücht. Züchtungsbiol. 84, 110–114.

Fonseca, W., 1977. O Búfalo: sinônimo de carne, leite, manteiga e trabalho. Ministério da Agricultura DNPA-DAGE e Associação Brasileira de Criadores de Búfalos, São Paulo.

Foulds, L.R., Penny, D., Hendy, M.D., 1979. A general approach to proving the minimality of phylogenetic trees illustrated by an example with a set of 23 vertebrates. J. Mol. Evol. 13, 151–166.

Giuffra, E., Kijas, J.M., Amarger, V., Carlborg, O., Jeon, J.T., Andersson, L., 2000. The origin of the domestic pig: independent domestication and subsequent introgression. Genetics 154, 1785–1791.

Harpending, H.C., 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. Hum. Biol. 66, 591–600.

Hasegawa, M., Kishino, H., 1989. Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders. Jpn. J. Genet. 64, 243–258.

Hassanin, A., Douzery, E.J., 1999a. Evolutionary affinities of the enigmatic saola (*Pseudoryx nghetinhensis*) in the context of the molecular phylogeny of Bovidae. Proc. R. Soc. Lond. B 266, 893–900.

Hassanin, A., Douzery, E.J., 1999b. The tribal radiation of the family Bovidae (Artiodactyla) and the evolution of the mitochondrial cytochrome *b* gene. Mol. Phylogenet. Evol. 13, 227–243.

Hiendleder, S., Mainz, K., Plante, Y., Lewalski, H., 1998. Analysis of mitochondrial DNA indicates that domestic sheep are derived from two different ancestral maternal sources: no evidence for contributions from urial and argali sheep. J. Hered. 89, 113–120.

Hudson, R.R., Kaplan, N.L., 1995. The coalescent process and background selection. Philos. Trans. R. Soc. Lond. B 349, 19–23.

Iannuzzi, L., 1994. Standard karyotype of the river buffalo (*Bubalus bubalis* L., $2n = 50$). Report of the committee for the standardization of banded karyotypes of the river buffalo. Cytogenet. Cell Genet. 67, 102–113.

Lau, C.H., Drinkwater, R.D., Yusoff, K., Tan, S.G., Hetzel, D.J., Barker, J.S., 1998. Genetic diversity of Asian water buffalo (*Bubalus bubalis*): mitochondrial DNA D-loop and cytochrome *b* sequence variation. Anim. Genet. 29, 253–264.

Loftus, R.T., Ertugrul, O., Harba, A.H., El-Barody, M.A., MacHugh, D.E., Park, S.D., Bradley, D.G., 1999. A microsatellite survey of cattle from a centre of origin: the Near East. Mol. Ecol. 8, 2015–2022.

Maddison, D.R., P., M.W., 2000. MacClade 4: Analysis of Phylogeny and Character Evolution, 4.0. Sinauer Associates, Sunderland, MA.

Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16, 1215.

Nachtsheim, H., Stengel, H., 1977. Vom Wildtier zum Haustier. Paul Parey Verlag, Berlin (West), Hamburg.

Nagata, J., Masuda, R., Tamate, H.B., Hamasaki, S., Ochiai, K., Asada, M., Tatsuzawa, S., Suda, K., Tado, H., Yoshida, M.C., 1999. Two genetically distinct lineages of the sika deer, *Cervus nippon*, in Japanese islands: comparison of mitochondrial D-loop region sequences. Mol. Phylogenet. Evol. 13, 511–519.

Nascimento, C.N.B., Carvalho, L.O.D.M., Lourenço Jr., J.B., 1979. Importância do búfalo para a pecuária brasileira. EMBRAPA/CPATU, Belém-PA, pp. 3–4.

Nei, M., 1987. Molecular Evolution Genetic. Columbia University Press, New York, USA.

Posada, D., Crandall, K.A., 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol. Evol. 16, 37–45.

Richards, M.B., Macaulay, V.A., Bandelt, H.J., Sykes, B.C., 1998. Phylogeography of mitochondrial DNA in western Europe. Ann. Hum. Genet. 62 (Pt. 3), 241–260.

Ritz, L.R., Glowatzki-Mullis, M.L., MacHugh, D.E., Gaillard, C., 2000. Phylogenetic analysis of the tribe Bovini using microsatellites. Anim. Genet. 31, 178–185.

Rogers, A.R., Harpending, H., 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9, 552–569.

Saccone, C., Attimonelli, M., Sbisa, E., 1987. Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. J. Mol. Evol. 26, 205–211.

Salerno, A., 1974. The buffaloes of Italy. In: Cockrill, W.R. (Ed.), The Husbandry and Health of the Domestic Buffalo. FAO, Rome, pp. 737–747.

Savage, D., Russell, D., 1983. Mammalian Paleofaunas of the World. Addison-Wesley, Reading, MA.

Sbisa, E., Tanzariello, F., Reyes, A., Pesole, G., Saccone, C., 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205, 125–140.

Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18, 502–504.

Schneider, S., Roessli, D., Excoffier, L., 2000. Arlequin: A Software Population Genetics Data Analysis, ver. 2.000. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Slatkin, M., Hudson, R.R., 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129, 555–562.

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.

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. Biochem. Genet. 34, 443–452.

Tanaka, K., Yamagata, T., Masangkay, J.S., Faruque, M.O., Vu-Binh, D., Salundik, Mansjoer, S.S., Kawamoto, Y., Namikawa, T., 1995. Nucleotide diversity of mitochondrial DNAs between the swamp and the river types of domestic water buffaloes, *Bubalus bubalis*, based on restriction endonuclease cleavage patterns. Biochem. Genet. 33, 137–148.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C., Bradley, D.G., 2001. Genetic evidence for Near-Eastern origins of European cattle. Nature 410, 1088–1091.

Ulbrich, F., Fischer, H., 1967. The chromosomes of the Asiatic buffalo (*Bubalus bubalis*) and the Afrikan buffalo (*Syncerus caffer*). Z. Tierzücht. Züchtungsbiol. 83, 219–223.

Watterson, G.A., Guess, H.A., 1977. Is the most frequent allele the oldest. Theor. Popul. Biol. 11, 141–160.

Wood, N.J., Phua, S.H., 1996. Variation in the control region sequence of the sheep mitochondrial genome. Anim. Genet. 27, 25–33.

Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39, 306–314.

Yang, Z., 1996. Phylogenetic analysis using parsimony and likelihood methods. J. Mol. Evol. 42, 294–307.