Molecular Phylogenetics and Evolution 161 (2021) 107170

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

The genome of the lowland anoa (*Bubalus depressicornis*) illuminates the origin of river and swamp buffalo

Manon Curaudeau [a], Roberto Rozzi [b,c], Alexandre Hassanin [a,*]

[a] Institut Systématique Evolution Biodiversité (ISYEB), Sorbonne Université, MNHN, CNRS, EPHE, UA, 57 rue Cuvier, CP 51, 75005 Paris, France
[b] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Synthesis Centre for Biodiversity Sciences (sDiv), Puschstr. 4, D-04103 Leipzig, Germany
[c] Museum für Naturkunde, Leibniz-Institut für Evolutions- und Biodiversitätsforschung, 10115 Berlin, Germany

ARTICLE INFO

Keywords:
Bubalus bubalis
River buffalo
Swamp buffalo
Domestication
Species delimitation
Pleistocene

ABSTRACT

Two types of domestic water buffalo are currently recognized: the river buffalo from the Indian subcontinent and Mediterranean countries and the swamp buffalo from China and Southeast Asia. To test the hypothesis of two separate species of water buffalo, we sequenced the genome of the lowland anoa, *Bubalus depressicornis*, which is a dwarf wild buffalo endemic to Sulawesi, and two genomes of swamp buffalo, and made comparisons with 12 additional genomes. Three genomic data sets were constructed to infer phylogenetic relationships: the mitochondrial genome (15,468 bp; maternal transmission), two concatenated Y-chromosomal genes, *AMELY* and *DDX3Y* (20,036 bp; paternal transmission), and a selection of 30 nuclear genes representing all cattle chromosomes (364,887 bp; biparental transmission). The comparisons between our 30 nuclear gene sequences obtained by read mapping and those directly extracted from *Bos taurus* and *Bubalus bubalis* genome assemblies show that the mapping approach revealed higher levels of heterozygosity at both nucleotide sites and indels (insertions and deletions) (0.09–0.15%), as well as several sequence errors (0.07%). Our phylogenetic and molecular dating analyses provide strong evidence that the lowland anoa, river buffalo, and swamp buffalo are three distinct taxa which separated rapidly from each other during the Pleistocene epoch. We therefore conclude that two species of domestic water buffalo should be distinguished: *Bubalus bubalis* for the river buffalo and *Bubalus kerabau* for the swamp buffalo. The new classification can have deep implications for understanding the evolution and selection of domesticated forms and for the conservation and management of wild buffalo populations in South and Southeast Asia.

1. Introduction

Buffaloes belong to the family Bovidae (Mammalia, Cetartiodactyla) in which they are currently classified in the Bubalina, a subtribe represented by two genera, *Bubalus* and *Syncerus* (Hassanin, 2014; IUCN, 2020). The genus *Syncerus* contains only the African buffalo, *Syncerus caffer* (Sparrman, 1779), a species endemic to Sub-Saharan Africa. The genus *Bubalus* includes four wild species, all found in tropical Asia: the wild water buffalo, *Bubalus arnee* (Kerr, 1792), which was widely distributed in India and Southeast Asia, is now restricted to a few populations in Bhutan, Cambodia, India, Nepal, and Thailand; the lowland anoa, *Bubalus depressicornis* (Smith, 1827), and the mountain anoa, *Bubalus quarlesi* (Ouwens, 1910), are both found on Sulawesi and Buton Island; and the tamaraw, *Bubalus mindorensis* Heude, 1888, is endemic to the island of Mindoro in the Philippines.

The domestic water buffalo, which represents a global population of 202 million, was originally found in the Indian subcontinent, China and Southeast Asia (Zhang et al., 2020). Several importations have been documented: in Europe (Italy) and Africa (via Egypt) during the Middle Age, and in Australia and South America during the 19th and 20th centuries. The domestic water buffalo is generally placed in its own species, *Bubalus bubalis* (Linnæus, 1758), separated from its putative wild progenitor, *B. arnee* (Gentry et al., 2004). Two types of domestic water buffalo are currently recognized: the river buffalo, which is mainly used for milk, is found in the Indian subcontinent and Mediterranean countries; and the swamp buffalo, which is primarily used as draft animal, occurs in China and Southeast Asia (Zhang et al., 2020). These two types differ in their body coloration (the river buffalo is black, whereas the swamp buffalo is usually dark grey with white chevrons on the throat and white socks) and their horns (curved in the river buffalo,

* Corresponding author.
E-mail address: alexandre.hassanin@mnhn.fr (A. Hassanin).

https://doi.org/10.1016/j.ympev.2021.107170
Received 17 December 2020; Received in revised form 12 March 2021; Accepted 25 March 2021
Available online 30 March 2021

straighter in the swamp buffalo) (MacGregor, 1941; Castello, 2016). Several studies have shown that river and swamp buffaloes are genetically divergent based on various molecular data, such as mtDNA, Y-chromosome genes, microsatellites, and SNP (Kierstein et al., 2004; Kumar et al., 2007; Zhang et al., 2007; Yindee et al., 2010; Zhang et al., 2016; Colli et al., 2018; Ravi Kumar et al., 2020). Some of these studies have concluded that the two types of buffalo were domesticated independently and should be classified as distinct subspecies (Kumar et al., 2007; Yindee et al., 2010). However, most of these studies did not include any molecular data from wild species of *Bubalus*. Intriguingly, an analysis of mitogenomes showed that the lowland anoa (*B. depressicornis*), river buffalo and swamp buffalo represent three mtDNA lineages differing by 2.1–2.3%, suggesting they belong to three distinct species recently separated during the Pleistocene epoch (Hassanin et al., 2012). However, species delimitation based only on mitochondrial sequences are known to be potentially misleading because these molecules are transmitted maternally. In particular, female philopatry, which is a common behaviour in mammals, can result in the evolution of divergent mtDNA haplogroups between distant populations of the same species (Li and Kokko, 2019). In mammals, males disperse further and more frequently than females, which allows genetic exchanges of nuclear alleles between distant populations, preventing genetic drift and therefore a possible speciation event. Since sexual differences in dispersal behaviour can result in discordant phylogenetic patterns between mitochondrial and nuclear genes (Hassanin et al., 2015; Petzold and Hassanin, 2020), the hypothesis that the lowland anoa, river buffalo and swamp buffalo are three closely related species needs to be tested using nuclear DNA sequences. Another problem with using mtDNA for species delimitation is that mitochondrial introgression is known to occur frequently between closely related species of mammals (Hassanin and Ropiquet 2007; Petzold and Hassanin, 2020).

In the present study, we developed a genomic approach for studying species delimitation within the genus *Bubalus* by analyzing 15 genomes, including three newly sequenced genomes from one lowland anoa and two swamp buffaloes (Table 1). Three genomic data sets representing the three modes of genetic inheritance (transmission) encountered in mammals were constructed and analyzed to compare phylogenetic patterns: the mitogenome (data set named *mtDNA*) for the maternal inheritance, two Y chromosome genes (*AMELY* and *DDBY;* data set named *Y*) for the paternal inheritance, and a selection of 30 genes extracted from each of the 30 cattle chromosomes (29 autosomes and X chromosome) (data set named *29A/X*) for the biparental inheritance. By sampling large genomic character sets and considering all modes of inheritance, this strategy should overcome problems due to rapid and recent diversification and should detect possible hybridization events (e.g. Kutschera et al., 2014; Li et al., 2016).

2. Material and methods

2.1. DNA extraction, genome sequencing and assembly

DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) from skin biopsies sampled on males of the lowland anoa (ANT3, animal named Yannick housed in captivity at the *Ménagerie du Jardin des Plantes*, MNHN) and two swamp buffaloes (V7x54, Vietnam; CKM7, Cambodia). The biopsies on living animals were made by a veterinary surgeon and all experimental protocols were approved by the Muséum national d'Histoire naturelle. DNA samples were quantified with a Qubit® 2.0 Fluorometer using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA).

The libraries and DNA sequencing were done at the "Institut du Cerveau et de la Moelle épinière": ANT3 was sequenced on a NextSeq® 500 Illumina system using the NextSeq 500 High Output Kit v2 (300 cycles), whereas V7x54 and CKM7 were sequenced on a Novaseq6000 Illumina system using the NovaSeq 6000 SP Reagent Kit (500 cycles) (Illumina, San Diego, CA, USA).

The three new genomes were assembled using the MaSuRCA assembler version 3.3.1 (Zimin et al., 2013) using the recommended parameters for mammalian genomes. The quality of the assemblies was estimated using Quast 2.2 (Gurevich et al., 2013).

2.2. Taxonomic sample

The Sequence Read Archives (SRA) available at the National Center for Biotechnology Information (NCBI) were used to retrieve high-quality Bovini genomes. We selected only Illumina paired-end DNA reads because of their low error rate (<2%; Nagarajan and Pop, 2013). Male individuals were preferred in order to extract sequences from the Y chromosome. We considered only shotgun sequencing with more than 30 Gbp to extract the selection of genes listed below with a coverage higher than 15X.

Fifteen individuals were used for the genomic analyses representing seven species of Bovidae: *Bos frontalis*, *Bos indicus* (two breeds), *Bos taurus* (two breeds), *B. bubalis*, including three river buffalo genomes and our two swamp buffalo genomes, *B. depressicornis*, *Capra hircus*, and *S. caffer* (Table 1). A species of the family Cervidae, *Odocoileus virginianus*, was used as outgroup.

2.3. Construction of the three genomic data sets

Three genomic data sets were constructed for the phylogeny of the Bovini: (1) the complete mitochondrial genome (data set named *mtDNA*), two Y chromosome genes (*AMELY* and *DDBY;* data set named *Y*) and a selection of 30 genes extracted from each of the 30

Table 1
Genomic data analyzed in this study.

Taxon	Country	Breed	Sex	Biosample
Bos indicus	India	Gir	Male	SAMN08225763*
Bos indicus	India	Hariana	Female	SAMEA5577149*
Bos frontalis	India	Nagaland	Female	SAMN02689702 (Mukherjee et al., 2019)
Bos taurus	England	Hereford	Female	SAMN03145444 (Rosen et al., 2020)
Bos taurus	Netherlands	Holstein	Male	SAMEA4780322
Bubalus bubalis	Bangladesh	Bangladeshi (river)	Male	SAMN05990785 (Mintoo et al., 2019)
Bubalus bubalis	India	Jafarabadi (river)	Male	SAMN00004269
Bubalus bubalis	Italy	Mediterranean (river)	Female	SAMN08640746 (Low et al., 2019)
Bubalus bubalis	Cambodia	NA (swamp)	Male	CKM7*
Bubalus bubalis	Vietnam	NA (swamp)	Male	V7x54*
Bubalus depressicornis	Indonesia	NA	Male	ANT3*
Capra hircus	Bangladesh	Black Bengal	Male	SAMN10460883
Capra hircus	Switzerland	Toggenburg	Male	SAMN12871781*
Odocoileus virginianus	United States	NA	Male	SAMN05363940*
Syncerus caffer	South Africa	NA	Male	SAMN05717674 (Glanzmann et al., 2016)

* Unpublished.

chromosomes (29 autosomes and X chromosome) of *B. taurus* and *C. hircus* (data set named *29A/X*).

The 30 genes of the *29A/X* data set were chosen with the following criteria: orthologous genes of 10–15 kb with conserved synteny between *B. taurus* and *C. hircus* and differing by less than 10% of their size in cattle and goat genomes. Using the Ensembl Genes 96 database available at http://www.ensembl.org/, we extracted the list of orthologous genes (orthology confidence = 1) in common between cattle (*B. taurus*, ARS-USD1.2) and goat genomes (*C. hircus*, ARS1), with their gene ID, chromosome number, chromosome position (gene start and end) and gene size. Then, we filtered the 1687 genes of 10–30 kb in length in both species that differ in size by less than 10% between cattle and goat genomes, and that show the same name and same chromosome number in both species. For each of the 30 chromosomes (29 autosomes + X chromosome), we then selected the five smallest genes >10 kb. The 150 genes extracted from the cattle genome were studied by BLAST search (ncbi-blast-2.9.0+; Altschul et al., 1990) against *C. hircus*. Finally, we focused on orthologous genes for which we found no more than two BLAST hits in *C. hircus* representing a total query coverage >90%, and selected 30 orthologous genes for phylogenetic analyses representing one gene of 10–15 kb for each of the 30 chromosomes of both *B. taurus* and *C. hircus* (Table 2).

To better account for heterozygosity, as well as possible assembly errors (as discussed in Section 3.1), the raw reads were mapped onto the genomic fragments extracted from five reference assemblies at the genus level (i.e. *Bos*, *Bubalus*, *Capra*, *Syncerus* and *Odocoileus*). For this purpose, the 33 genomic fragments detailed in Table 2 were extracted from the ARS-UCD1.2 assembly of *B. taurus* (Rosen et al., 2020) and they were blasted (BLAST 2.9.0; Altschul et al., 1990) against the five following assemblies representing different genera: *B. taurus* (NCBI accession number: GCA_002263795.2) for *Bos*, *C. hircus* (GCA_001704415.1) for *Capra*, *B. depressicornis* (this study) for *Bubalus*, *S. caffer* (GCA_006408785.1) for *Syncerus*, and *O. virginianus* (GCA_002102435.1) for *Odocoileus*. BLAST outputs were concatenated in case of multiple hits on the same fragment, extended by 10 kb on 5′ and 3′ sides (to account for large indels), realigned and trimmed with respect to the *B. taurus* reference sequence. The SRA of each of the six *Bubalus* samples (Table 1) were then mapped onto the 33 genomic fragments of *B. depressicornis* using Geneious Prime 2019.2.3 (https://www.geneious.com) with a mismatch of 3%. Similarly, the SRA of each of the five *Bos* samples (Table 1) were mapped onto the 33 genomic fragments of *B. taurus*. Due to possible differences with the genome assembly references, in particular large indels or assembly errors, the mapping may have gaps. To fill them, the consensus sequences of the contigs were used as new references for a second mapping using a mismatch of 2%. To detect heterozygous sites, the consensus sequences of final contigs were constructed using a threshold set to 75% (bases called in the consensus match at least 75% of the reads).

Table 2
Thirty-three genomic fragments analyzed in this study.

Markers	Chromosomes*	Sizes (bp)*	Substitution models
AMBN	6	11 650	GTR + G
APEH	22	10 884	HKY + I
CDAN1	10	12 266	GTR + I
CSTA	1	12 549	GTR + G
CYP2E1	26	11 282	GTR + G
FKBP3	21	11 157	GTR + I
GSDMB	19	11 484	GTR + G
HBEFG	7	11 329	HKY + G
HPD	17	12 286	GTR + G
IDO1	27	15 067	GTR + G
KDELC1	12	14 269	GTR + G
KRT79	5	11 973	GTR + G
LRRC32	15	12 274	GTR + G
MICAL1	9	12 475	HKY + G
NDUFS2	3	10 439	GTR + G
NMS	11	10 719	GTR + G
NOL6	8	13 163	GTR + G
NUPL2	4	16 046	GTR + G
PRMT1	18	10 264	HKY + G
PTGER4	20	13 084	GTR + G
RER1	16	10 913	GTR + I
RNF40	25	12 187	GTR + G
SCYL1	29	11 308	GTR + G
SLC11A1	2	11 738	GTR + G
TAPBP	23	13 243	GTR + G
TFE3	X	11 756	GTR + G
TONSL	14	11 297	GTR + I
TRMT6	13	10 628	GTR + G
TXNL4A	24	11 722	GTR + G
ZSWIM8	28	15 435	GTR + G
Y (*AMELY* + *DDBY*)	Y	20 036	GTR + G
mtDNA	Mitogenome	15 468	GTR + I + G

* in *Bos taurus* and *Capra hircus*.

The 33 genomic fragments extracted for 15 taxa were aligned in Geneious 2019.2.3 with MAFFT version 7.450 (Katoh and Standley, 2013) using default parameters, and the alignments were corrected manually using the three following criteria: (1) transitions were privileged over transversions because they are more frequent; (2) the number of indels was minimized because they are rarer events than nucleotide substitutions; (3) gaps were placed in 3′ position for alignment reproducibility. The mitochondrial DNA control region was removed because its alignment was too ambiguous for primary homology. The *Y* data set was constructed by concatenating the two Y chromosome genes. The *29A/X* data set was constructed by concatenating the 29 autosomal genes and the X chromosome gene. The 33 genomic alignments analyzed in this study are available at https://osf.io/46m2e/.

2.4. Phylogenetic analyses

The three genomic data sets (*mtDNA*, *Y*, *29A/X*) and the 30 independent genes of the *29A/*X data set (Table 2) were analyzed for phylogenetic reconstruction. The best-fit nucleotide substitution model was selected for each of the 33 genomic fragments (*mtDNA*, two Y chromosome genes, and the 30 genes of the *29A/X data set*) by using the Akaike information criterion (AIC) in ModelTest-NG (Darriba et al., 2020). The selected models are listed in Table 2.

Bayesian analyses were performed with MrBayes v3.2.7 (Ronquist et al., 2012) on the three genomic data sets (*mtDNA*, *Y*, and *29A/X*) and a specific model was used for each independent marker of the two concatenated data sets, *29A/X* and *Y* (partitioned approach). Each of the 30 independent genes of the *29A/X* data set were also analyzed separately. The posterior probabilities were computed with four independent Markov chains for ten million generations with a sampling every 1000 generations and a burn-in of 25%.

Bootstrap percentages were calculated for the three genomic data sets (*mtDNA*, *Y*, *29A/X*) using the maximum likelihood approach under PhyML 3.0 (Guindon et al., 2010), 1000 bootstrap replicates, a starting tree inferred with BIONJ, and the BEST algorithm for tree rearrangement (best of SPR – Subtree Pruning and Regrafting, and NNI – Nearest Neighbor Interchanges).

The lists of bipartitions obtained from the separate Bayesian analyses of the 32 independent markers (*mtDNA*, *Y*, and the 30 genes of the *29A/X* data set) were used as inputs in SuperTRI v.57 (available at http://www.normalesup.org/~bli/Programs/programs.html) to determine the levels of congruence between the 32 data sets (Ropiquet et al., 2009). With the SuperTRI method, three reliability values were calculated for each node of interest: the mean posterior probability (MPP), the index of reproducibility (Rep), which is the ratio of the number of data sets supporting the node of interest to the total number of data sets, and the SuperTRI bootstrap percentage (SBP). All values were reported on the SuperTRI bootstrap majority-rule consensus tree reconstructed from the weighted parsimony analysis of the MRP (Matrix Representation with Parsimony) matrix of 597 binary characters (automatically coded using SuperTRI v57) using PAUP 4* version 4b10 (Swofford, 2003) after

1000 bootstrap replicates (addition sequence: closest).

The uncorrected pairwise distances (p–distance) were computed using PAUP* version 4.0a167 (Swofford, 2003) for each of three genomic data sets (*mtDNA*, *Y*, *29A/X*).

2.5. Molecular dating

Divergence times were estimated on the CIPRES Science Gateway (Miller et al., 2010) using the concatenation of the three genomic data sets (*mtDNA* + *Y* + *29A/X*) and the Bayesian approach implemented in BEAST v.2.4.7 (Bouckaert et al., 2014). A GTR + I + G model was used for each of the three data sets. Four calibration points were used for molecular dating with a normal distribution: the most recent common ancestor (MRCA) of Ruminantia was set at 23.4 ± 2.6 Mya and that of Bovidae at 19.7 ± 1.9 Mya in agreement with Hassanin et al. (2012); the MRCA of Bubalina was set at 7.95 ± 2.25 Mya in agreement with the fossil record (minimum age = 5.7 Mya, *Ugandax* (Gentry, 2010); maximum age = 10.2 Mya, *Selenoportax* (Bibi, 2007)); and the MRCA of *B. taurus* and *B. indicus* was set at 0.305 ± 0.030 Mya in agreement with Bradley et al. (1996) and Achilli et al. (2009). Node ages were estimated using a calibrated Yule speciation prior and 3×10^8 generations, with tree sampling every 1000 generations, and a burn-in of 10^7 generations. MCMC mixing efficiency and convergence were assessed using the ESS values (>200) in Tracer v.1.7.1 (Rambaut et al., 2018). The chronogram was reconstructed with TreeAnnotator, which is included in the BEAST package (Bouckaert et al., 2014).

3. Results and discussion

3.1. Consensus sequences generated by reads mapping versus genome assembly

Nuclear genes usually show a small amount of nucleotide variation at the genus, species, and subspecies levels. Some nucleotide positions, called heterozygous sites, can be variable because of differences in the two parental alleles. Divergent alleles can be the signature of two significant phenomena: incomplete lineage sorting or interspecific introgression. However, heterozygosity is known to be underestimated with most methods of de novo assembly (Kajitani et al., 2019). To better account for heterozygosity, as well as possible assembly errors, we constructed consensus sequences (threshold set to 75% for consensus base calling) for a selection of 30 nuclear genes and 15 taxa by mapping Illumina reads to the reference genes extracted from either *B. taurus* or *B. bubalis* genome assemblies, ARS-UCD1.2 (Rosen et al., 2020) and UOA_WB_1 (Low et al., 2019), respectively. We found many differences between our gene sequences and those directly extracted from genome assembly (Table 3; Appendix A): 313 differences in *B. taurus* sequences (detected in 23/30 nuclear genes) and 655 differences in *B. bubalis* sequences (detected in 29/30 nuclear genes). As expected, most of the differences concern heterozygous sites (82.7% for *B. taurus* and 79.5% for *B. bubalis*), but we also found a significant amount of indels (15.0% for *B. taurus* and 12.1% for *B. bubalis*) and nucleotide errors (2.2% for *B. taurus* and 8.4% for *B. bubalis*). Using the assembly sequences directly would result in an underestimation of heterozygosity representing 0.09% and 0.15% in the 30 nuclear genes of *B. taurus* and *B. bubalis*, respectively, as well as 0.07% of uncorrected errors in the 30 nuclear genes of both *B. taurus* and *B. bubalis*. The impact of these differences on phylogenetic analyses among closely related species or subspecies may

Table 3
Differences between the nuclear sequences extracted directly from the genome assemblies of *Bos taurus* (Hereford breed) and *Bubalus bubalis* (Mediterranean breed) and those obtained after read mapping (the pairwise comparisons are detailed in Appendix A).

	Bos taurus						*Bubalus bubalis*					
Genes	Diff (%)	H sites	H indels	Sub	Del	In	Diff (%)	H sites	H indels	Sub	Del	In
AMBN	0	0	0	0	0	0	0.19	13	0	1	2	1
APEH	0.14	13	1	0	0	0	0.27	14	0	6	2	4
CDAN1	0.01	1	0	0	0	0	0.03	3	0	0	1	0
CSTA	0.05	5	0	0	0	1	0.24	23	1	0	1	1
CYP2E1	0.08	9	0	0	0	0	0.04	0	1	0	0	0
FKBP3	0	0	0	0	0	0	0.17	12	1	0	1	1
GSDMB	0.1	9	0	0	0	2	0.35	27	0	4	1	1
HBEFG	0.04	4	0	0	0	0	0.2	18	0	2	1	0
HPD	0.01	0	0	0	0	1	0.24	18	0	5	0	4
IDO1	0.46	50	0	2	3	6	0.46	48	0	2	0	6
KDELC1	0	0	0	0	0	0	0.1	5	0	0	1	0
KRT79	0	0	0	0	0	0	0.47	40	0	0	2	1
LRRC32	0.01	1	0	0	0	0	0.25	21	0	2	1	3
MICAL1	0	0	0	0	0	0	0.2	18	0	2	2	1
NDUFS2	0.01	1	0	0	0	0	0.47	38	0	3	1	2
NMS	0	0	0	0	0	0	0.66	44	0	3	2	5
NOL6	0.12	14	0	1	0	0	0.22	22	0	1	0	1
NUPL2	0.99	78	2	1	2	10	0.33	1	0	2	1	0
PRMT1	0.09	7	0	0	1	1	0.14	9	0	3	0	1
PTGER4	0.34	29	1	0	2	5	0.38	25	0	2	3	2
RER1	0.06	6	0	0	0	0	0.37	28	1	6	1	4
RNF40	0.21	18	1	0	1	1	0.02	1	0	1	0	0
SCYL1	0.05	2	1	0	1	1	0.15	14	0	1	0	1
SLC11A1	0.01	0	0	1	0	0	0.04	4	0	0	0	0
TAPBP	0.02	2	0	0	0	0	0.12	6	0	3	1	1
TFE3	1.61*	2	0	2	1	1	0.08	5	0	1	2	1
TONSL	0	0	0	0	0	0	0.13	10	0	1	3	0
TRMT6	0.04	4	0	0	0	0	0	0	0	0	0	0
TXNL4A	0.03	2	0	0	0	1	0.39	33	0	3	1	2
ZSWIM8	0.01	2	0	0	0	0	0.16	21	0	1	2	0
Total	0.16	259	6	7	11	30	0.23	521	4	55	32	43

Abbreviations: Del = deletions; Diff = differences; H = heterozygous; indels = insertions-deletions; In = Insertions; Sub = substitutions.
* high value explained by a large indel of 182 bp.

be highly misleading as our *29A/X* data set contains very small proportions of variable and informative sites at the generic levels: 0.57% and 0.40% of variable sites for *Bos* and *Bubalus*, respectively; and 0.12% and 0.13% of informative sites for *Bos* and *Bubalus*, respectively.

3.2. The river buffalo and swamp buffalo belong to two different taxa

The phylogenetic analyses based on either the mitogenome (*mtDNA*, Fig. 1), the two Y chromosome genes (*Y*, Fig. 1), or the 30 independent genes representing the 30 bovid chromosomes (*29A/X*, Fig. 2) show maximal support values (posterior probability [PP] = 1; bootstrap percentage [BP] = 100) for the monophyly of the tribe Bovini (cattle and buffalo), the two subtribes Bovina (*Bos*) and Bubalina (*Bubalus* + *Syncerus*), the genus *Bubalus*, the species *B. taurus* (cattle) and *B. indicus* (zebu), as well as for the sister-group relationship between *B. taurus* and *B. indicus*. These results agree with previous molecular studies on the systematics of the tribe Bovini (Hassanin and Ropiquet, 2004; MacEachern et al., 2009; Hassanin et al., 2013).

The monophyly of the river buffalo and that of the swamp buffalo are also supported by maximal values (PP = 1; BP = 100) with the three genomic data sets (*mtDNA*, *Y*, and *29A/X*). The SuperTRI analyses also indicate that most separate analyses of the 32 independent markers support the monophyly of the river buffalo (SBP = 98; MPP = 0.38; Rep = 0.44) and that of the swamp buffalo (SBP = 100; MPP = 0.73; Rep = 0.84). These results confirm with strong support some previous genetic findings inferred from mtDNA sequences (Kierstein et al., 2004; Kumar et al., 2007; Lei et al., 2007; Zhang et al., 2016), Y chromosome sequences (Yindee et al., 2010; Zhang et al., 2016), and autosomal markers, such as allozymes (Barker et al., 1997), microsatellites (Barker et al., 1997; Zhang et al., 2011), and SNP markers (Colli et al., 2018; Luo et al., 2020; Ravi Kumar et al., 2020). Given the substantial genetic differences between river and swamp buffalo, several authors argued that they should be classified as distinct subspecies (Kumar et al., 2007; Luo et al., 2020).

By including the lowland anoa (*B. depressicornis*) in the genomic comparisons, our analyses show however that its phylogenetic position is unstable with respect to the two buffalo lineages: it appears as the sister-group of *B. bubalis* in the *29A/X* tree (PP = 0.97 and BP = 54) (Fig. 2A); whereas it is more closely related to the swamp buffalo in the *Y* tree (PP = 0.35 and BP = 56) (Fig. 1B), SuperTRI bootstrap consensus tree (SBP = 98; MPP = 0.38; Rep = 0.41) (Fig. 2B) and BEAST tree based on the concatenation of the three data sets (PP = 1) (Fig. 3); and it is more closely related to the river buffalo in the *mtDNA* tree (PP = 0.79 and BP = 50) (Fig. 1A). Since none of the three possible phylogenetic hypotheses is highly supported (BP ≤ 56), interrelationships between the three *Bubalus* taxa can be considered as unresolved. Obviously, this is the consequence of a rapid diversification (as discussed below). Taxonomically, these results are consistent with two possible interpretations: either the anoa should be regarded as a subspecies of *B. bubalis*, or, alternatively, the two types of domestic buffalo should be treated as separate species. The pairwise distances calculated from the three genomic data sets (Table 4) are comparable between the lowland anoa, river buffalo and swamp buffalo: 0.18–0.22% for *29A/X*; 0.17–0.20% for *Y*; and 2.19–2.37% for *mtDNA*. These distances are similar or slightly more important than interspecific distances calculated between cattle (*B. taurus*) and zebu (*B. indicus*): 0.15–0.17% for *29A/X*; 0.15–0.17% for *Y*; and 1.26–1.28% for *mtDNA* (Table 4), suggesting that the lowland anoa, river buffalo and swamp buffalo should be considered as three closely related species of the genus *Bubalus*. This taxonomic view is also supported by morphological data. The lowland anoa is a

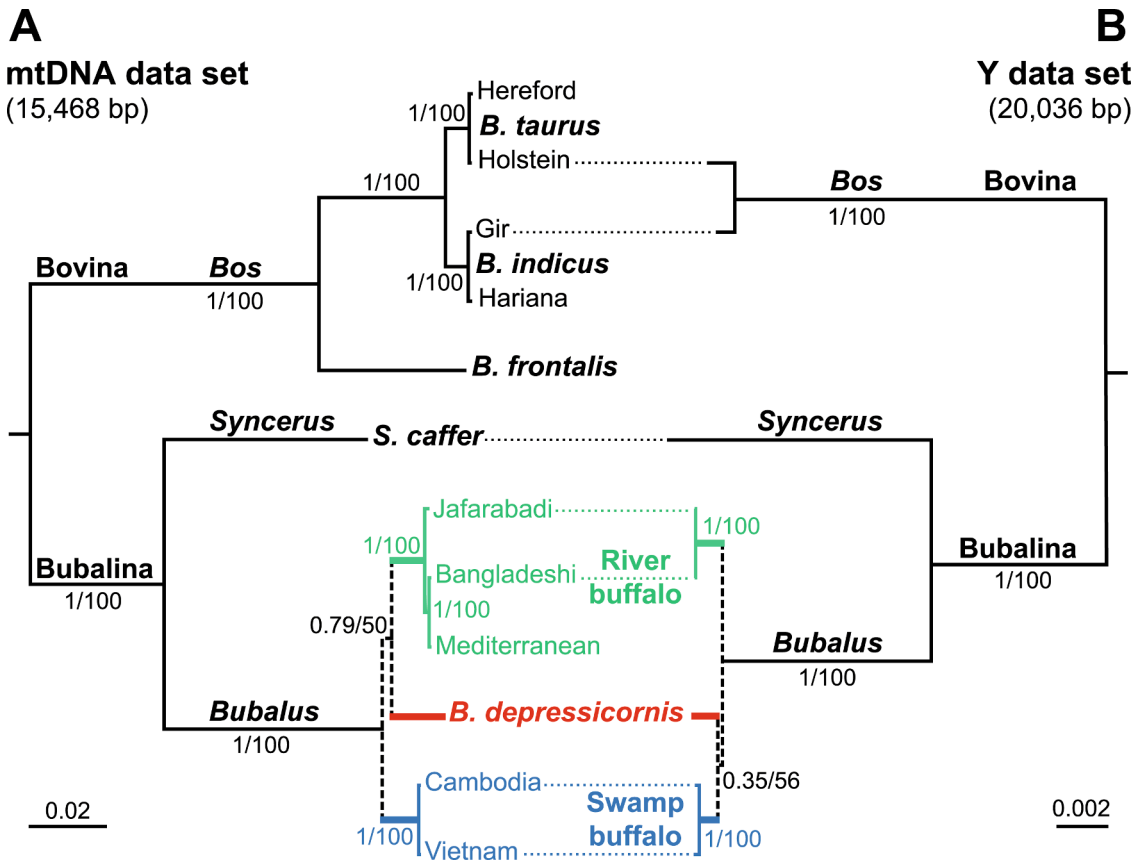

Fig. 1. Bayesian trees based on the mitogenome (A; *mtDNA* data set) and the concatenation of two Y chromosome genes, *AMELY* and *DBY* (B; *Y* data set). The two values on the branches are the Bayesian posterior probability (on the left) and the Maximum likelihood bootstrap percentage (on the right). Dashed branches indicate unresolved relationships.

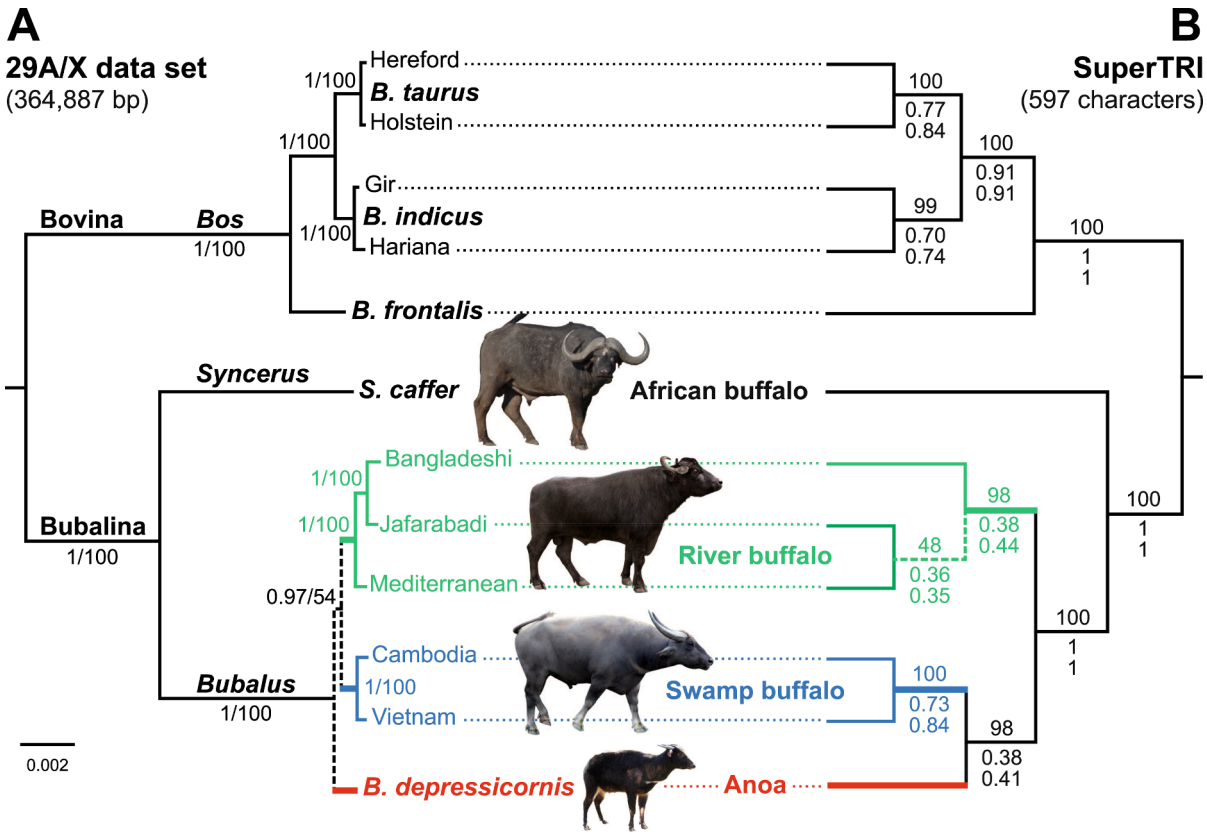

Fig. 2. Bayesian tree based on the concatenation of the 30 genes of the *29A/X* data set (A) and the SuperTRI bootstrap majority-rule consensus tree (B). A. The two values on the branches are the Bayesian posterior probability (on the left) and the Maximum likelihood bootstrap percentage (on the right). B. The three SuperTRI values on the branches are top to bottom: the SuperTRI bootstrap percentage (SBP), the mean posterior probability (MPP), and the index of reproducibility (Rep). Dashed branches indicate unresolved relationships. The photos of river buffalo and swamp buffalo are from Castello (2016) and those of African buffalo and lowland anoa (named Yannick, sequenced in this study) are from AH.

Table 4
Pairwise distances calculated from the three genomic data sets.

Taxa (number of individuals)	Distances		
	29A/X	*Y*	*mtDNA*
B. taurus intraspecific (n = 2)	0.02	NA	0.06
B. indicus intraspecific (n = 2)	0.04	NA	0.11
B. bubalis intraspecific (n = 3)	0.08 ± 0.01	0	0.11 ± 0.08
B. kerabau intraspecific (n = 2)	0.06	0.02	0.03
B. bubalis versus B. kerabau (n = 5)	0.17 ± 0.02	0.18 ± 0.00	2.02 ± 0.01
B. bubalis versus B. depressicornis (n = 4)	0.20 ± 0.02	0.20 ± 0.00	2.20 ± 0.01
B. kerabau versus B. depressicornis (n = 3)	0.18 ± 0.00	0.17 ± 0.00	2.37 ± 0.00
B. taurus versus B. indicus (n = 4)	0.16 ± 0.01	0.17	1.27 ± 0.01

wild dwarf buffalo found only on Sulawesi and Buton Island, where it lives in solitary in lowland forests and wetlands (Burton et al., 2016). It is morphologically clearly distinct from all other species of the tribe Bovini (see Fig. 2) based on body size (shoulder height < 0.9 m; weight < 300 kg), horn shape (straight horns pointed backwards, not curved and only slightly divergent), horn length (males: 271–373 mm; females: 183–260 mm) and coat coloration (the general colour is black with white legs and often a white crescent on the throat) (Groves, 1969; Rozzi, 2017). It was therefore included in its own (sub)genus *Anoa* in some previous classifications (Groves, 1969; Rozzi, 2017). The two types of domestic water buffalo have distinct body coloration: the river buffalo is black, whereas the swamp buffalo is usually dark grey with white chevrons on the throat and white socks (MacGregor, 1941; Castello, 2016). In addition, the horns of the river buffalo show a double curvature (at first, they are directed downward and backward, and then curl upward in a spiral), whereas those of the swamp buffalo are semi-circular and always remain approximately in the same plane as the forehead (MacGregor, 1941; Castello, 2016). Their species status is also supported by cytogenetic evidence: the river buffalo has 2n = 50 chromosomes and FN = 58 (Iannuzzi, 1994), whereas the lowland anoa and swamp buffalo share the same diploid number (2n = 48) but differ in the number of major chromosome arms (termed fundamental number, FN), i.e., FN = 58 in the lowland anoa and FN = 56 in the swamp buffalo (Nguyen et al., 2008). Despite their different karyotypes, river and swamp buffaloes can produce viable and fertile F1 hybrids, with 2n = 49 (Degrandi et al., 2014). However, spermatozoa abnormalities are significantly more frequent in male hybrids than in river and swamp bulls (Dai et al., 1994).

3.3. What should be the names of the two species of domestic buffalo?

The river buffalo has been domesticated in the western region of the Indian subcontinent ca. 6300 years BP. From about the seventh century, the river buffalo was introduced in Italy and south-eastern Europe, where it was used as a draft and dairy animal (Zhang et al., 2016; Zhang et al., 2020). Thomas (1911) found that the species described by Linnaeus (1758) as *B. bubalis* refers to the domestic buffalo, and fixed the type locality as "Italy (Rome)". Thus, the name *B. bubalis* should be used for the river buffalo. The swamp buffalo has been domesticated in the China/Indochina border region ca. 3000–7000 years BP, and then spread south through peninsular Malaysia to the Sunda Islands

(Sumatra, Java and Sulawesi), north into central China, and then through an eastern island route via Taiwan to the Philippines and Borneo (Zhang et al., 2020). In recent publications, the swamp buffalo was often named *Bubalus bubalis carabanensis* following Castillo (1998), who mentioned that its common name is *carabao* in the Philippines, *karbaw* in Indonesia, *kerbau sawak* in Malaysia, and *kwai* in Thailand. A few years before, Castillo (1971) treated the swamp buffalo as a full species, *Bubalus carabanensis*. However, *B. carabanensis* Castillo, 1971 is a junior synonym of *Bubalus kerabau* Fitzinger, 1860 (Wilson and Reeder, 2005; Castello, 2016). Thus, the scientific name of the swamp buffalo should be *B. kerabau* in accordance with the principle of priority (article 23 of the International Code of Zoological Nomenclature; https://www.iczn.org/the-code/).

3.4. Radiation of buffalo species in the late Early Pleistocene

Our divergence time estimates and their 95% confidence intervals (CI) (Fig. 3) suggest that *Bubalus* diverged from *Syncerus* during the Late Miocene, at around 6.3 Mya (95% CI: 8.55–4.13 Mya), and that its diversification into three species corresponding to *B. depressicornis*, *B. bubalis* and *B. kerabau* occurred at around 0.84 Mya (95% CI: 1.28–0.49 Mya). Such period fits well with molecular estimates published for the common ancestor of river and swamp buffalo in previous studies, 1.3 ± 0.5 Mya (Hassanin et al., 2012) and 0.9–0.86 Mya (Hassanin and Ropiquet, 2004). A high diversity of buffalo species was described in the Middle and Late Pleistocene of China, Europe, and Southeast Asia, and some fossils were dated from the Early Pleistocene (Masini et al., 2013; Rozzi et al., 2013; Dong et al., 2014; Filoux et al., 2015; Fauzi et al., 2016; Rozzi, 2017; Koenigswald et al., 2019). The most recent common ancestor of *B. bubalis*, *B. depressicornis*, and *B. kerabau* was likely to disperse in India, Southeast Asian mainland and Sunda Islands in the late Early Pleistocene when the cooling accelerated from 1.2 to 0.9 Mya (McClymont et al., 2013). During glacial periods, the sea level can drop up to 120 m below present-day sea level, allowing faunal dispersals from Southeast Asian mainland to India, or reciprocally, through the coasts in the Bay of Bengal and from Southeast Asian mainland to Sunda Islands (Meijaard, 2004). In agreement with this scenario, Rozzi (2017) proposed that the most reasonable ancestor of the lowland and mountain anoas would be *Bubalus palaeokerabau*, occurring on Java since the Early/Middle Pleistocene (van den Bergh et al., 2001; Rozzi, 2018; Volmer et al., 2019).

3.5. Conclusion on the wild water buffalo

According to the International Commission on Zoological Nomenclature (Gentry et al., 2004), wild forms of water buffalo should be named *B. arnee* (Kerr, 1792). The species is in danger of extinction: it is believed to be extinct in Bangladesh, Malaysia, and on the islands of Sumatra, Java, and Borneo; and there are just a few remnant populations in southern Nepal, southern Bhutan, western Thailand, eastern Cambodia, northern Myanmar, and several sites in Central India and North-east India (Kaul and Williams, 2019). Based on the geographic distribution of river and swamp buffalo in South Asia and Southeast Asia, respectively, it can be inferred that the species *B. bubalis* has been domesticated in India, whereas *B. kerabau* has been domesticated in northern Indochina (Zhang et al., 2020). Morphologically, the populations of *B. arnee* show important variations, in particular in coat colour and horn shape, although their appearances are more similar to the swamp buffalo (Groves, 1996). Genetic data on wild water buffalo are scarce but they confirm the West-East dichotomy detected in domestic lineages: based on 18 microsatellite loci, a Nepal wild buffalo from the Kosi Tappu wildlife reserve appeared as the sister-group of river buffalo breeds (Zhang et al., 2011); based on mtDNA sequences of the control region, a Thai wild buffalo from the Huai Kha Khaeng wildlife sanctuary clustered as the sister-group of haplogroup SA (Sarataphan et al., 2017), one of the five major mitochondrial haplogroups detected in swamp buffalo breeds (Wang et al., 2017; haplogroup named SBa in Youssef et al., 2021). For conservation purpose, it is therefore urgent to sequence the genomes of wild buffaloes previously included in the three subspecies recognized by Groves (1996) based on differences

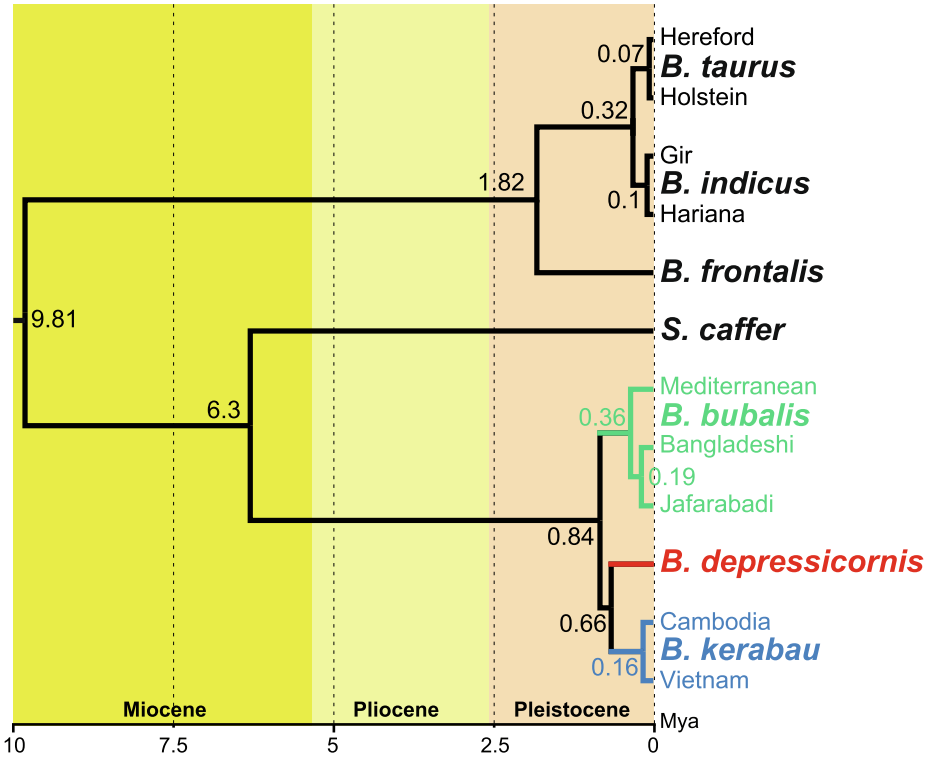

Fig. 3. BEAST chronogram inferred from the concatenation of the three genomic data sets. The mean divergence times are reported on the nodes and the horizontal grey bars show 95% confidence intervals. All nodes were supported by PP = 1.

in body size, skull dimensions, coat coloration and horn shape: *Bubalus arnee arnee* in central India (Madhya Pradesh) and Nepal, *Bubalus arnee fulvus* in the Brahmaputra valley, and *Bubalus arnee theerapati* in Thailand, Cambodia and Vietnam. Importantly, the buffaloes of the Brahmaputra valley might belong to a distinct species, or alternatively the Brahmaputra valley might be a hybrid zone between two species.

Funding

This work was supported by the Alexander Von Humboldt Foundation and a grant from the "Agence Nationale de la Recherche" under the LabEx ANR-10-LABX-0003-BCDiv in the program "Investissements d'avenir" (ANR-11-IDEX-0004-02). R.R. was supported by sDiv, Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (DFG– FZT 118, 202548816), by the Alexander Von Humboldt Foundation, and by the German Research Foundation (DFG RO 5835/2-1).

CRediT authorship contribution statement

Manon Curaudeau: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization. **Roberto Rozzi:** Resources, Writing - review & editing, Funding acquisition. **Alexandre Hassanin:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the people who helped to collect tissue samples used in this study: senator Nhim Vanda, Jean-Luc Berthier, Norin Chai, Gerard Dousseau, Claire Rejaud, Trung Thanh Nguyen, Marie-Lilith Patou, and Do Tuoc. AH acknowledges José R. Castello who provided the permission to use the pictures of river buffalo and swamp buffalo published in his book. We also thank the two anonymous reviewers for their helpful comments on the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107170.

References

Achilli, A., Bonfiglio, S., Olivieri, A., Malusa, A., Pala, M., Kashani, B.H., Perego, U.A., Ajmone-Marsan, P., Liotta, L., Semino, O., Bandelt, H.J., Ferretti, L., Torroni, A., 2009. The multifaceted origin of taurine cattle reflected by the mitochondrial genome. PloS One 4, e5753. https://doi.org/10.1371/journal.pone.0005753.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Barker, J.S.F., Moore, S.S., Hetzel, D.J.S., Evans, D., Tan, S.G., Byrne, K., 1997. Genetic diversity of Asian water buffalo (*Bubalus bubalis*): microsatellite variation and a comparison with protein-coning loci. Anim. Genet. 28, 103–115. https://doi.org/10.1111/j.1365-2052.1997.00085.x.

Bibi, F., 2007. Origin, paleoecology, and paleobiogeography of early Bovini. Palaeogeogr. Palaeoclimatol. Palaeoecol. 248, 60–72. https://doi.org/10.1016/j.palaeo.2006.11.009.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537 https://doi.org/10.1371/journal.pcbi.1003537.

Bradley, D.G., MacHugh, D.E., Cunningham, P., Loftus, R.T., 1996. Mitochondrial diversity and the origins of African and European cattle. Proc. Natl. Acad. Sci. 93, 5131–5135. https://doi.org/10.1073/pnas.93.10.5131.

Burton, J., Wheeler, P., Mustari, A., 2016. Bubalus depressicornis. The IUCN Red List of Threatened Species 2016, e.T3126A46364222. Downloaded on 10 July 2020. doi: 10.2305/IUCN.UK.2016-2.RLTS.T3126A46364222.

Castello, J.R., 2016. Bovids of the World Antelopes, Gazelles, Cattle, Goats, Sheep, and Relatives. Princeton University Press.

Castillo, L.S., 1971. Proposal for a new scientific name for the carabao – *Bubalus carabanensis* (Linn.). Philippine J. Animal Sci. 8, 155–156.

Castillo, L.S., 1998 Proposal: new scientific name of the domesticated swamp buffalo, the carabao – Bubalus bubalis carabanensis [(Sub) Sp. Nov. Castillo 1998]. In: 20. Annual Scientific Meeting of the National Academy of Science and Technology, Manila (Philippines), 8–9 Jul 1998.

Colli, L., Milanesi, M., Vajana, E., Iamartino, D., Bomba, L., Puglisi, F., Corvo, M.D., Nicolazzi, E.L., Ahmed, S.S.E., Herrera, J.R.V., et al., 2018. New insights on water buffalo genomic diversity and post-domestication migration routes from medium density SNP chip data. Front. Genet. 9, 53. https://doi.org/10.3389/fgene.2018.00053.

Dai, K., Gillies, C.B., Dollin, A.E., Hilmi, M., 1994. Synaptonemal complex analysis of hybrid and purebred water buffaloes (*Bubalus bubalis*). Hereditas 121, 171–184. https://doi.org/10.1111/j.1601-5223.1994.00171.x. PMID: 7876032.

Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T., 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol. Biol. Evol. 37, 291–294. https://doi.org/10.1093/molbev/msz189.

Degrandi, T.M., Pita, S., Panzera, Y., de Oliveira, E.H., Marques, J.R., Figueiró, M.R., Marques, L.C., Vinadé, L., Gunski, R.J., Garnero, A.V., 2014. Karyotypic evolution of ribosomal sites in buffalo subspecies and their crossbreed. Genet. Mol. Biol. 37, 375–380. https://doi.org/10.1590/s1415-47572014000300009.

Dong, W., Liu, J.-y., Zhang, L.-m., Xu, Q.-q., 2014. The Early Pleistocene water buffalo associated with Gigantopithecus from Chongzuo in southern China. Quat. Int. 354, 86–93. https://doi.org/10.1016/j.quaint.2013.12.054.

Fauzi, M.R., Ansyori, M., Prastiningtyas, D., Intan, M., Wibowo, U., Wulandari, H., Rahmanendra, H., Widianto, H., Simanjuntak, T., et al., 2016. Matar: A forgotten but promising Pleistocene locality in East Java. Quat. Int. 416, 183–192. https://doi.org/10.1016/j.quaint.2015.12.091.

Filoux, A., Wattanapituksakul, A., Lespes, C., Thongcharoenchaikit, C., 2015. A Pleistocene mammal assemblage containing Ailuropoda and Pongo from Tham Prakai Phet cave, Chaiyaphum Province, Thailand. Geobios 48, 341–349. https://doi.org/10.1016/j.geobios.2015.07.003.

Fitzinger, L.J., 1860. Der Sunda-Büffel (Bubalus Kerabau). Wissenschaftlich-populäre Naturgeschichte der Säugethiere in ihren sämmtlichen Hauptformen, V. Kaiserlich-Königlichen Hof- und Staatsdruckerei, Wien, p. 329.

Gentry, A., Clutton-Brock, J., Groves, C.P., 2004. The naming of wild animal species and their domestic derivatives. J. Archaeol. Sci. 31, 645–651. https://doi.org/10.1016/j.jas.2003.10.006.

Gentry, A., 2010. Bovidae. In: Werdelini, L., Sanders, W.J. (eds.), Cenozoic Mammals of Africa. University of California Press, Berkeley. pp.741–796.

Glanzmann, B., Möller, M., Le Roex, N., Tromp, G., Hoal, E.G., Van Helden, P.D., 2016. The complete genome sequence of the African buffalo (*Syncerus caffer*). BMC Genomics 17, 1001. https://doi.org/10.1186/s12864-016-3364-0.

Groves, C.P., 1969. Systematics of the anoa (Mammalia, Bovidae). Beaufortia 17, 1–12.

Groves, C.P., 1996. The taxonomy of the Asian Wild Buffalo from the Asian mainland. Zeitschrift fuer Saeugetierkunde 61, 327–338.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321. https://doi.org/10.1093/sysbio/syq010.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

Hassanin, A., An, J., Ropiquet, A., Nguyen, T.T., Couloux, A., 2013. Combining multiple autosomal introns for studying shallow phylogeny and taxonomy of Laurasiatherian mammals: Application to the tribe Bovini (Cetartiodactyla, Bovidae). Mol. Phylogenet. Evol. 66, 766–775. https://doi.org/10.1016/j.ympev.2012.11.003.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., van Vuuren, B.J., Matthee, C., Ruiz-Garcia, M., Catzeflis, F., Areskoug, V., Nguyen, T.T., et al., 2012. Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. C.R. Biol. 335, 32–50. https://doi.org/10.1016/j.crvi.2011.11.002.

Hassanin, A., Khouider, S., Gembu, G.C., Goodman, S.M., Kadjo, B., Nesi, N., Pourrut, X., Nakouné, E., Bonillo, C., 2015. The comparative phylogeography of fruit bats of the tribe scotonycterini (chiroptera, pteropodidae) reveals cryptic species diversity related to african pleistocene forest refugia. C.R. Biol. 338, 197–211. https://doi.org/10.1016/j.crvi.2014.12.003.

Hassanin, A., Ropiquet, A., 2004. Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Kouprey, *Bos sauveli* Urbain 1937. Mol. Phylogenet. Evol. 33, 896–907. https://doi.org/10.1016/j.ympev.2004.08.009.

Hassanin, A., Ropiquet, A., 2007. Resolving a zoological mystery: the kouprey is a real species. Proc. Biol. Sci. 274, 2849–2855. https://doi.org/10.1098/rspb.2007.0830.

Hassanin, A., 2014. Systematic and evolution of Bovini. In: Melletti, M., Burton, J. (eds). Ecology, Evolution and Behaviour of Wild Cattle: Implications for Conservation. Cambridge University Press, pp. 7–20.

Heude, P.M., 1888. Note sur le petit buffle sauvage de l'ile de Mindoro (Philippines). Mémoires concernant l'histoire naturelle de l'Empire chinois 2 (4), 50.

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.
IUCN, 2020. The IUCN Red List of Threatened Species. Version 2020-2. Downloaded on 09 July 2020. Retrieved from https://www.iucnredlist.org.
Kajitani, R., Yoshimura, D., Okuno, M., Minakuchi, Y., Kagoshima, H., Fujiyama, A., Kubokawa, K., Kohara, Y., Toyoda, A., Itoh, T., 2019. Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. Nat. Commun. 10 (1), 1702. https://doi.org/10.1038/s41467-019-09575-2.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.
Kaul, R., Williams, A., rithe, k., Steinmetz, R., Mishra, R., 2019 Bubalus arnee. The IUCN Red List of Threatened Species 2019, e.T3129A46364616. Downloaded on 10 July 2020. doi:10.2305/IUCN.UK.2019-1.RLTS.T3129A46364616.en.
Kerr, R., 1792. R.. Arnee Bos arnee. The Animal Kingdom or zoological system of the celebrated Sir Charles Linnaeus. Class I. Mammalia. Strahan A. and Cadell T., Edinburgh & London, p. 336.
Kierstein, G., Vallinoto, M., Silva, A., Schneider, M.P., Iannuzzi, L., Brenig, B., 2004. Analysis of mitochondrial d-loop region casts new light on domestic water buffalo (*Bubalus bubalis*) phylogeny. Mol. Phylogenet. Evol. 30, 308–324. https://doi.org/10.1016/s1055-7903(03)00221-5.
Koenigswald, W.v., Schwermann, A.H., Keiter, M., Menger, F., 2019. First evidence of Pleistocene *Bubalus murrensis* in France and the stratigraphic occurrences of Bubalus in Europe. Quaternary Int. 522, 85–93. https://doi.org/10.1016/j.quaint.2019.06.019.
Kumar, S., Nagarajan, M., Sandhu, J.S., Kumar, N., Behl, V., Nishanth, G., 2007. Mitochondrial dna analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo. Animal Genet. 38, 227–232. https://doi.org/10.1111/j.1365-2052.2007.01602.x.
Kutschera, V.E., Bidon, T., Hailer, F., Rodi, J.L., Fain, S.R., Janke, A., 2014. Bears in a forest of gene trees: Phylogenetic inference is complicated by incomplete lineage sorting and gene flow. Mol. Biol. Evol. 31, 2004–2017. https://doi.org/10.1093/molbev/msu186.
Lei, C., Zhang, W., Chen, H., Lu, F., Liu, R., Yang, X., Zhang, H., Liu, Z., Yao, L., Lu, Z., Zhao, Z.L., 2007. Independent maternal origin of Chinese swamp buffalo (*Bubalus bubalis*). Animal Genet. 38, 97–102. https://doi.org/10.1111/j.1365-2052.2007.01567.x.
Li, G., Davis, B.W., Eizirik, E., Murphy, W.J., 2016. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). Genome Res. 26, 1–11. https://doi.org/10.1101/gr.186668.114.
Li, X.-Y., Kokko, H., 2019. Sex-biased dispersal: A review of the theory. Biol. Rev. 94, 721–736. https://doi.org/10.1111/brv.12475.
Linnæus C. 1758 Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I. Holmiæ.
Low, W.Y., Tearle, R., Bickhart, D.M., Rosen, B.D., Kingan, S.B., Swale, T., Thibaud-Nissen, F., Murphy, T.D., Young, R., Lefevre, L., et al., 2019. Chromosome-level assembly of the water buffalo genome surpasses hu- man and goat genomes in sequence conti-guity. Nature Commun. 10, 260. https://doi.org/10.1038/s41467-018-08260-0.
Luo, X., Zhou, Y., Zhang, B., Zhang, Y., Wang, X., Feng, T., Li, Z., Cui, K., Wang, Z., Luo, C., et al., 2020. Understanding divergent domestication traits from the whole-genome sequencing of swamp and river-buffalo populations. Natl. Sci. Rev. 7, 686–701. https://doi.org/10.1093/nsr/nwaa024.
MacEachern, S., McEwan, J., Goddard, M., 2009. Phylogenetic reconstruction and the identification of ancient polymorphism in the Bovini tribe (Bovidae, Bovinae). BMC Genomics 10, 177. https://doi.org/10.1186/1471-2164-10-177.
MacGregor, R., 1941. The domestic buffalo. Vet. Rec. 53, 443–450.
Masini, F., Palombo, M.R., Rozzi, R., 2013. A reappraisal of the early to middle pleistocene Italian bovidae. Quat. Int. 288, 45–62. https://doi.org/10.1016/j.quaint.2012.03.026.
McClymont, E.L., Sosdian, S.M., Rosell-Melé, A., Rosenthal, Y., 2013. Pleistocene sea-surface temperature evolution: Early cooling, delayed glacial intensification, and implications for the mid-Pleistocene climate transition. Earth Sci. Rev. 123, 173–193. https://doi.org/10.1016/j.earscirev.2013.04.006.
Meijaard, E., 2004. Solving mammalian riddles: a reconstruction of the Tertiary and Quaternary distribution of mammals and their palaeoenvironments in island South-East Asia. PhD thesis. School of Archaeology and Anthropology The Australian National University, Canberra, Australia.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 gateway computing environments workshop (GCE). Ieee, pp. 1–8.
Mintoo, A.A., Zhang, H., Chen, C., Moniruzzaman, M., Deng, T., Anam, M., Emdadul Huque, Q.M., Guang, X., Wang, P., Zhong, Z., et al., 2019. Draft genome of the river water buffalo. Ecol. Evol. 9, 3378–3388. https://doi.org/10.1002/ece3.4965.
Mukherjee, S., Cai, Z., Mukherjee, A., Longkumer, I., Mech, M., Vupru, K., Khate, K., Rajkhowa, C., Mitra, A., Guldbrandtsen, B., et al., 2019. Whole genome sequence and de novo assembly revealed genomic architecture of Indian Mithun (*Bos frontalis*). BMC Genomics 20, 617. https://doi.org/10.1186/s12864-019-5980-y.
Nagarajan, N., Pop, M., 2013. Sequence assembly demystified. Nat. Rev. Genet. 14, 157–167. https://doi.org/10.1038/nrg3367.
Nguyen, T.T., Aniskin, V.M., Gerbault-Seureau, M., Planton, H., Renard, J.P., Nguyen, B. X., Hassanin, A., Volobouev, V.T., 2008. Phylogenetic position of the saola (*Pseudoryx nghetinhensis*) inferred from cytogenetic analysis of eleven species of Bovidae. Cytogenetic Genome Res. 122, 41–54. https://doi.org/10.1159/000151315.
Ouwens, P.A., 1910. Contribution a la connaissance des mammiféres de Célébès. Bull. Dépt. Agric. Indes Néerl. 38 (Zool., 6), 1–7.
Petzold, A., Hassanin, A., 2020. A comparative approach for species delimitation based on multiple methods of multi-locus DNA sequence analysis: A case study of the genus *Giraffa* (Mammalia, Cetartiodactyla). PLoS ONE 15, e0217956. https://doi.org/10.1371/journal.pone.0217956.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901. https://doi.org/10.1093/sysbio/syy032.
Ravi Kumar, D., Devadasan, M.J., Surya, T., Vineeth, M.R., Choudhary, A., Sivalingam, J., KatariaRS, N.SK., Tantia, M.S., Verma, A., 2020. Genomic diversity and selection sweeps identified in Indian swamp buffaloes reveals it's uniqueness with riverine buffaloes. Genomics 112, 2385–2392. https://doi.org/10.1016/j.ygeno.2020.01.010.
Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. https://doi.org/10.1093/sysbio/sys029.
Ropiquet, A., Li, B., Hassanin, A., 2009. SuperTRI: A new approach based on branch support analyses of multiple independent data sets for assessing reliability of phylogenetic inferences. C.R. Biol. 332, 832–847. https://doi.org/10.1016/j.crvi.2009.05.001.
Rosen, B.D., Bickhart, D.M., Schnabel, R.D., Koren, S., Elsik, C.G., Tseng, E., Rowan, T.N., Low, W.Y., Zimin, A., Couldrey, C., et al., 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. GigaScience 9, giaa021. doi: 10.1093/gigascience/giaa021.
Rozzi, R., Winkler, D.E., De Vos, J., Schulz, E., Palombo, M.R., 2013. The enigmatic bovid *Duboisia santeng* (Dubois, 1891) from the Early-Middle Pleistocene of Java: A multiproxy approach to its paleoecology. Palaeogeogr. Palaeoclimatol. Palaeoecol. 377, 73–85. https://doi.org/10.1016/j.palaeo.2013.03.012.
Rozzi, R., 2017. A new extinct dwarfed buffalo from Sulawesi and the evolution of the subgenus *Anoa*: an interdisciplinary perspective. Quat. Sci. Rev. 157, 188–205. https://doi.org/10.1016/j.quascirev.2016.12.011.
Rozzi, R., 2018. Space-time patterns of body mass variation in island bovids: The key role of predatory release. J. Biogeogr. 45, 1196–1207. https://doi.org/10.1111/jbi.13197.
Sarataphan, N., Narongwanichgarn, W., Maneerat, S., 2017. Phylogenetic analysis of a Thai wild water buffalo (*Bubalus arnee*) through mitochondrial control region. Int. J. Conserv. Sci. 8, 105–112.
Smith, C.H., 1827. The seventh order of the Mammalia. The Ruminantia. In: Griffith, E., Smith, C.H., Pidgeon, E. (Eds.), The animal kingdom arranged in conformity with its organization, by the Baron Cuvier, member of the Institute of France, with additional descriptions of all the species hitherto named, and of many not before noticed. Whittaker G.B., London, p. 293.
Sparrman, A., 1779. Bos Caffer, et nytt Species af Buffel, från Caput Bonæ Spei. Kongliga Vetenskaps Academiens Handlingar 40 (1–3), 79–84.
Swofford, D.L., 2003. PAUP*: phylogenetic analysis using parsimony, version 4.0a167.
Thomas, O., 1911. The Mammals of the Tenth Edition of Linnæeus; an Attempt to fix the Types of the Genera and the exact Bases and Localities of the Species. Proc. Zool. Soc. London 81, 120–158. doi:10.1111/j.1469-7998.1911.tb06995.x.
van den Bergh, G.D., de Vos, J., Sondaar, P.Y., 2001. The Late Quaternary palaeogeography of mammal evolution in the Indonesian Archipelago. Palaeogeogr. Palaeoclimatol. Palaeoecol. 171, 385–408. https://doi.org/10.1016/S0031-0182(01)00255-3.
Volmer, R., van der Geer, A.A., Cabrera, P.A., Wibowo, U.P., Kurniawan, I., 2019. When did Cuon reach Java?–Reinvestigation of canid fossils from Homo erectus faunas. Geobios 55, 89–102. https://doi.org/10.1016/j.geobios.2019.06.004.
Wang, S., Chen, N., Capodiferro, M.R., Zhang, T., Lancioni, H., Zhang, H., Miao, Y., Chanthakhoun, V., Wanapat, M., Yindee, M., et al., 2017. Whole mitogenomes reveal the history of swamp buffalo: initially shaped by glacial periods and eventually modelled by domestication. Sci. Rep. 7, 4708. https://doi.org/10.1038/s41598-017-04830-2.
Wilson, D.E., Reeder, D.M., 2005. Mammal species of the world: a taxonomic and geographic reference, vol. 1. Johns Hopkins University Press.
Yindee, M., Vlamings, B.H., Wajjwalku, W., Techakumphu, M., Lohachit, C., Sirivaidyapong, S., Thitaram, C., Amarasinghe, A.A.A.W.K., Alexander, P.A.B.D.A., Colenbrander, B., et al., 2010. Y-chromosomal variation confirms independent domestications of swamp and river buffalo. Anim. Genet. 41, 433–435. https://doi.org/10.1111/j.1365-2052.2010.02020.x.
Youssef, N.A., Curaudeau, M., El Nahas, S.M., Hassan, A., Hassanin, A., 2021. Haplotype diversity in the mitochondrial genome of the Egyptian river buffalo (*Bubalus bubalis*). *Mitochondrial DNA B Resour* 6, 145–147. https://doi.org/10.1080/23802359.2020.1852622.
Zhang, Y., Colli, L., Barker, J., 2020. Asian water buffalo: domestication, history and genetics. Anim. Genet. 51 (2), 177–191. https://doi.org/10.1111/age.12911.
Zhang, Y., Lu, Y., Yindee, M., Li, K.Y., Kuo, H.Y., Ju, Y.T., Ye, S., Faruque, M.O., Li, Q., Wang, Y., et al., 2016. Strong and stable geographic differentiation of swamp buffalo maternal and paternal lineages indicates domestication in the china/indochina border region. Mol. Ecol. 25, 1530–1550. https://doi.org/10.1111/mec.13518.

Zhang, Y., Sun, D., Yu, Y., Zhang, Y., 2007. Genetic diversity and differentiation of chinese domestic buffalo based on 30 microsatellite markers. Anim. Genet. 38, 569–575. https://doi.org/10.1111/j.1365-2052.2007.01648.x.

Zhang, Y., Vankan, D., Zhang, Y., Barker, J.S., 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. Anim. Genet. 42, 366–377. https://doi.org/10.1111/j.1365-2052.2010.02166.x.

Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., Yorke, J.A., 2013. The masurca genome assembler. Bioinformatics 29, 2669–2677. https://doi.org/10.1093/bioinformatics/btt476.