

## DNA barcoding of 18 species of Bovidae

CAI YanSen<sup>1,3</sup>, ZHANG Liang<sup>2\*</sup>, SHEN FuJun<sup>2</sup>, ZHANG WenPing<sup>2</sup>, HOU Rong<sup>2</sup>,  
YUE BiSong<sup>3</sup>, LI Jing<sup>3</sup> & ZHANG ZhiHe<sup>2\*</sup>

<sup>1</sup> Department of Medical Biology and Genetics, Luzhou Medical College, Luzhou 646000, China;

<sup>2</sup> Key Laboratory for Conservation Biology on Endangered Wildlife of Sichuan Province, Chengdu Research Base of Giant Panda Breeding, Chengdu 610000, China;

<sup>3</sup> Key Laboratory of Conservation Biology on Endangered Wildlife of Sichuan Province, College of Life Sciences, Sichuan University, Chengdu 610000, China

Received September 21, 2010; accepted November 26, 2010

Genetic divergences of mitochondrial cytochrome *c* oxidase subunit I genes, known as DNA barcodes, have been used in species identification in the animal kingdom. Barcodes can assist field workers and taxonomists to determine groups in need of taxa analysis, and facilitate the recognition of appropriate populations and scales for conservation planning. In this study, 18 species of Bovidae were selected to evaluate the effectiveness of DNA barcoding for species differentiation. The results showed that all but 2 species had unique DNA barcodes. The mean intraspecific variation was 0.63%, yielding a threshold of 6.3% for flagging putative species. The results supported the inference that barcode variation within species of mammals is somewhat higher than within other animal groups. The present study validated the effectiveness of barcoding for the identification of bovid species.

**DNA barcode, COI, Bovidae, conservation, endangered species**

**Citation:** Cai Y S, Zhang L, Shen F J, et al. DNA barcoding of 18 species of Bovidae. Chinese Sci Bull, 2011, 56, doi: 10.1007/s11434-010-4302-1

The advent of molecular techniques has opened up new possibilities for taxonomic research, which is important given that the vast majority of extant species are not morphologically well characterized [1]. Taxonomic assessments based on morphological analyses can be problematic due to phenotypic convergence among remotely related species or failure to distinguish ‘cryptic species’ where morphological divergence has not kept pace with genetic divergence [1].

Several studies suggest that a standardized 648-bp segment of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene may serve as a molecular marker for species identification in the animal kingdom [2–10]. For most animal groups, this gene region can be easily recovered. It provides fast and accurate species level discrimination, and allows discovery of new species across the tree of life [10]. Previous DNA barcode studies led to large-scale barcoding campaigns for various animal groups, such as birds, fish,

and lepidopterans [11]. Although DNA barcoding is increasingly regarded as efficient, existing work on mammals has been limited to a few studies of primates and small mammals [1,6,8,12,13].

However, some researchers, particularly taxonomists, are suspicious of DNA barcoding [14], especially with regard to the application of a universal distance criterion for species recognition [15]. In addition, critics have argued that a test for the precision of DNA barcoding should include a large proportion of closely related taxa [14,16,17].

In this study, we investigate the effectiveness of COI barcodes for species delineation among 223 individuals of Bovidae, representing 18 species. Bovidae is a mammalian family that is distributed worldwide. These large herbivores are not only particularly good tests for DNA barcoding, but also key species in the conservation of biodiversity. Many Chinese wild populations of Bovidae are currently at risk because of habitat loss and other threats, such as poaching and smuggling. These large mammals’ distribution and

\*Corresponding authors (email: zliang@panda.org.cn; zhangzh@mail.sc.cninfo.net)

taxonomy are better known than most other taxa, and rigorous planning of their protection will conserve key habitats for many other taxa. Therefore, elucidating their genetic relationships from small tissue fragments is particularly useful in monitoring biodiversity and illegal trades for endangered and threatened species. Furthermore, barcoding analysis of these organisms provides a molecular reference that can facilitate non-invasive sampling of hair, blood, or feces, which can be of high value in fieldwork and in captive population management.

## 1 Materials and methods

### 1.1 Taxon sampling

We extracted DNA from 45 individuals, representing 10 species of Bovidae. Among them, 7 species have been classified as state protected species in China (3 in level I and 4 in level II). We then retrieved 178 COI barcode sequences from 11 species from the Barcode of Life Data System (BOLD; <http://www.barcodinglife.com>). In total, 223 sequences from 18 species of Bovidae were analyzed.

Samples (muscle, liver tissue, blood, and hair follicles) were obtained from the Key Laboratory of Conservation Biology on Endangered Wildlife of Sichuan Province, Sichuan University, Chengdu Research Base of Giant Panda Breeding, and Chengdu Zoo, all in Chengdu, China. All specimens were preserved at  $-80^{\circ}\text{C}$ , either dry or in 95% ethanol. Collection localities and other information on the specimens are available in the Table S1.

### 1.2 PCR and sequencing

Total genomic DNA from fresh blood, hair follicles, and frozen tissues (muscle, liver, and blood) were extracted using standard protocols [18]. Total genomic DNA from feces samples were extracted using the QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany). The target region of the *COI* gene was amplified predominantly using 2 primer cocktails. Cocktail 1 was Zlf04 (5'-TCT CAA CTA AYC AYA AAG AYA TYG G-3') and Zlr04 (5'-TAA ACT TCR GGG TGA CCA AAR AAT CA-3'), while cocktail 2 [19] was VF1d (5'-TTC TCA ACC AAC CAC AAR GAY ATY GG-3') and VR1d (5'-TAG ACT TCT GGG TGG CCR AAR AAY CA-3'). The 25- $\mu\text{L}$  PCR reaction mix included 17  $\mu\text{L}$  of ultrapure water, 2.5  $\mu\text{L}$  of  $\text{MgCl}_2$ , 2.5  $\mu\text{L}$  of 10 $\times$  PCR buffer, 1  $\mu\text{L}$  of each primer (0.25 mmol/L), 0.25  $\mu\text{L}$  of each dNTP (0.05 mmol/L), 1.0 U of *Taq* polymerase, and 0.5–2  $\mu\text{L}$  of template DNA. The amplification protocol consisted of 5 min at  $94^{\circ}\text{C}$ ; followed by 35 cycles of 45 s at  $94^{\circ}\text{C}$ , 45 s at  $51^{\circ}\text{C}$ , and 45 s at  $72^{\circ}\text{C}$ ; and a final extension of 10 min at  $72^{\circ}\text{C}$ . PCR products were visualized on a 1% (w/v) agarose gel. PCR reactions generating a single product of about 700 bp were sequenced. Gel purification was used to recover the target fragment in cases where more

than one band was present: the recovered fragments were also sequenced. Sequencing reactions were carried out using Big Dye v3.1 and bidirectional PCR primers, and the products were analyzed on an ABI 377 sequencer (USA). The nucleotide sequence data generated are available in GenBank under accession numbers HQ269423–HQ269467.

### 1.3 Data analysis

Sequences containing insertions, deletions, nonsense or stop codons were considered as having arisen from PCR/sequencing errors or represented pseudogenes and were therefore excluded from the analyses. In addition to checking whether a sequence had derived from the true mitochondrial *COI* gene and not a numt (mitochondrial pseudogene in the nucleus), we determined whether the fragments contained any double peaks in the chromatogram, and if they were identical to, or at least were clustered with, orthologous mitochondrial sequences obtained from GenBank [20].

Sequence divergences among and within species were calculated using the Kimura-2-parameter (K2P) distance method in MEGA 3.1 [21]. Orthologous positions containing gaps were excluded from analyses using the 'complete deletion' option, and the vertebrate mitochondrial code was applied throughout. A Neighbor-Joining (NJ) tree [22] based on K2P distances was created to provide a graphic representation of the divergence patterns for among- and within-species. Node support was assessed by the bootstrap method [23] using 1000 pseudoreplicates (Figure 1, only bootstrap support values  $>70\%$  are shown).

## 2 Results

*COI* was amplified from all 45 individuals belonging to 10 species and 7 genera of the family Bovidae. For species represented by more than one individual in our study ( $n = 14$ ), *COI* sequences of the conspecific members were either identical or very similar to each other. However, among sequences acquired from BOLD, 4 individuals of *Bos taurus* were identical to members of *Bos indicus*. Hybridization between the 2 species is well known; therefore, we considered them to be hybrids and excluded them from further analyses (EU177868, EU177869, DQ124403, and EU177870). Apart from this exception, the other 16 species all had unique *COI* sequences and no barcodes were shared between species.

To estimate the efficiency of DNA barcodes in delimiting bovid species, we built a distance matrix including sequences generated by this study and those from BOLD (Table 1). The mean K2P sequence divergence within the 18 species of Bovidae examined in this study was 0.63%, while the mean divergence between congeners was about 7 $\times$  higher (4.89%). The mean divergence among species within



**Table 1** Genetic divergences of 223 COI sequences representing 18 species of Bovidae

Taxonomic levels	Minimum distance	Mean distance	Maximum distance
Species	0	0.63%	12.44%
Genus	1.41%	4.89%	8.34%
Family	2.37%	16.17%	20.17%

a family was 16.17%.

In most cases the NJ tree showed shallow intraspecific and deep interspecific divergences (Figure 1). However, extremely deep divergence was observed in one species, *Bubalus bubalis*, where the average and maximum intraspecific divergences were 7.09% and 12.44%.

### 3 Discussion

Several previous barcoding studies on vertebrates have raised concerns regarding the acquisition and ease of interpretation of DNA barcode data. One of the difficulties in amplifying the barcode region in vertebrates is caused by the use of universal primers [6,24]; however, this problem can often be solved using degenerated primers [6,25–28]. After encountering early difficulties in the barcode recovery for bovid species with the regular primer pair LCO1490 and HCO2198 [29], we turned to two other primer cocktails (cocktail 1: VF1d/VR1d, and cocktail 2: Zlf04/Zlr04). These primers contained degenerated positions, and primer cocktail 2 was designed specifically for bovid species, allowing more successful amplification. Applying this primer design strategy could result in higher success rates of barcode amplification, not only for Bovidae animals, but also for a broad range of animal taxa.

In addition to difficulties in PCR amplification for the barcode region, concerns have been raised on the potential influences of COI numts on barcoding analyses [30]. Previous barcode studies showed that numts were only detected in a very small percentage of bat species [6]. However, Song et al. [30] revealed that a large number of paralogous haplotypes with various divergences were coamplified with the orthologous mtDNA sequences when conserved primers were used for species with a large number of numts of the COI gene, and that the sequence divergences might overestimate the species numbers. In this study, we detected the presence of numts in only 2 fecal samples (1 *Budorcas taxicolor* and 1 *Connochaetes taurinus*), both of which clearly showed diagnostic mutations such as indels or stop codons. However, after we re-extracted DNA from hair follicles or blood, orthologous COI amplicons were obtained. Therefore, we strongly suggest that all barcoding researchers should check for the presence of pseudogenes, and that DNA extraction should be performed in mitochondrion-rich tissues.

Existing interpretations of DNA barcode data have shown that over 95% of animal species possess diagnostic

barcode sequences. For example, Ward et al. [3] found that all 207 Australian fish species included in their study had diagnostic barcode sequences, Hajibabaei et al. [4] found 98% of Costa Rican lepidopterans had diagnostic barcode arrays, and Kerr et al. [7] showed that about 94% of North American birds possessed distinct barcode clusters. In mammals, Clare et al. [6] and Borisenko et al. [8] revealed that barcodes enabled the discrimination of all species of Neotropical bats and small mammals that were examined in their studies. The present study reaffirmed the effectiveness of barcoding in large bovid herbivores. All but 2 of the bovid species examined possessed distinct COI sequences. Thirteen out of the 14 species are represented by multiple individuals, and the maximum intraspecific divergences in all but one species are lower than 0.8%. All 18 species possess minimum interspecific distances greater than 1.4%. The only exception was *Bubalus bubalis*, which showed mean and maximum intraspecific divergences of 7.09% and 12.44%.

However, the classification of *B. bubalis* is subject to debate. Whether it is a single species or 4 closely related, but separate, species is uncertain [31,32]. In previous studies, a ‘10× rule’, a threshold of 10 fold of the mean intraspecific variation, was proposed as a measure to screen for splits referred to as putative species [2,7,33]. In the present study, the results showed an average divergence within species of 0.63%, yielding a threshold of 6.3% to flag putative species. This implies that there probably are multiple species under the current concept of *B. bubalis*. Another way to flag putative species is to search for the conspecific groups whose specimens showed two or more distinct clusters with high bootstrap support in a NJ tree [33]. If this method was applied to our data, only *B. bubalis* would be flagged, and its species split would be supported.

It is noteworthy that intraspecific barcode variations in mammals are higher than those in other groups. For example, the intraspecific variation averaged 0.63% in bovid species, 1.1% in primates [1], 0.60% in Neotropical bats [6], and 1% in small mammal communities [8]. By contrast, the mean intraspecific variations in birds, fishes, and Lepidoptera were 0.27%, 0.39% and 0.46%, respectively. This elevated variation may represent the richness of diversity or could reflect some unique aspect of mitochondrial evolution [6]. However, the present test is preliminary. Further studies with more representatives of mammals and more individuals for each case should be performed to clarify the cause of such intraspecific genetic variation.

### 4 Conclusions

Our results suggest that DNA barcodes provide highly effective identification systems for bovid species. Depositing barcode sequences in a public database, along with primer sequences, trace files, and associated quality scores, will make

this species identification technique widely accessible [1]. The assembly of a DNA barcode library for mammals will not only aid species recognition, but will also lead to the development of an automated identification system, which would be particularly valuable for law enforcement and allow conservation officials to identify poachers and smugglers. However, the present study only investigated a small proportion of Bovidae species (about 12%), and our specimens were mainly collected in Sichuan province, China. For further studies, more comprehensive taxonomic samples, as well as populations from other geographical regions, are needed.

We thank CAI Guo, RAN JiangHong for field sampling; and ZHANG Xiu-Yue, MENG Yang, XU Xiao, WANG XiaoFang, ZENG Bo and LI YuZhi for their technical assistance. This work was supported by the grant from Chengdu Giant Panda Breeding Research Foundation (CPF08004).

- 1 Lorenz J G, Jackson W E, Beck J C, et al. The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philos Trans R Soc Lond B Biol Sci*, 2005, 360: 1869–1878
- 2 Hebert P D N, Stoeckle M Y, Zemlak T S, et al. Identification of birds through DNA Barcodes. *PLoS Biol*, 2004, 2: e312
- 3 Ward R D, Zemlak T S, Innes B H, et al. Barcoding Australia's fish species. *Philos Trans R Soc Lond B Biol Sci*, 2005, 360: 1847–1857
- 4 Hajibabaei M, Janzen D H, Burns J M, et al. DNA barcodes distinguish species of tropical Lepidoptera. *Proc Natl Acad Sci USA*, 2006, 103: 968–971
- 5 Yoo H S, Eah J Y, Kim J S, et al. DNA barcoding Korean birds. *Mol Cell*, 2006, 22: 323–327
- 6 Clare E L, Lim B K, Engstrom M D, et al. DNA barcoding of Neotropical bats: Species identification and discovery within Guyana. *Mol Ecol Notes*, 2007, 7: 184–190
- 7 Kerr K C R, Stoeckle M Y, Dove C J, et al. Comprehensive DNA barcode coverage of North American birds. *Mol Ecol Notes*, 2007, 7: 535–543
- 8 Borisenko A V, Lim B K, Ivanova N V, et al. DNA barcoding in surveys of small mammal communities: A field study in Suriname. *Mol Ecol Resour*, 2008, 8: 471–479
- 9 Tavares E S, Baker A J. Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. *BMC Evol Biol*, 2008, 8: 81
- 10 Hebert P D N, Cywinska A, Ball S L. Biological identifications through DNA barcodes. *Proc R Soc Lond B Biol Sci*, 2003, 270: 313–321
- 11 Marshall E. Will DNA barcodes breathe life into classification? *Science*, 2005, 307: 1037
- 12 Hajibabaei M, Singer G A C, Hickey D A. Benchmarking DNA barcodes: An assessment using available primate sequences. *Genome*, 2006, 49: 851–854
- 13 Robins J, Hingston M, Matisoo-Smith E, et al. Identifying *Rattus* species using mitochondrial DNA. *Mol Ecol Notes*, 2007, 7: 717–729
- 14 Wiemers M, Fiedler K. Does the DNA barcoding gap exist? A case study in blue butterflies (Lepidoptera: Lycaenidae). *Front Zool*, 2007, 4: 8
- 15 Pereira S L, Baker A J. A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Mol Biol Evol*, 2006, 23: 1731–1740
- 16 Moritz C, Cicero C. DNA barcoding: Promise and pitfalls. *PLoS Biol*, 2004, 2: e354
- 17 Meyer C P, Paulay G. DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol*, 2005, 3: e422
- 18 Sambrook J, Fritsch E F, Maniatis T. *Molecular Cloning, A Laboratory Manual*. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989
- 19 Ivanova N V, deWaard J R, Hebert P D N. An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Mol Ecol Notes*, 2006, 6: 998–1002
- 20 Hanner R. Proposed Standards for BARCODE Records in INSDC (BRIs). Database Working Group, Consortium for the Barcode of Life, 2005
- 21 Kimura M. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J Mol Evol*, 1980, 16: 111–120
- 22 Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 1987, 4: 406–425
- 23 Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 1985, 39: 783–791
- 24 Vences M, Thomas M, Bonett R M, et al. Deciphering amphibian diversity through DNA barcoding: Chances and challenges. *Philos Trans R Soc Lond B Biol Sci*, 2005, 360: 1859–1868
- 25 Batzer M A, Carlton J E, Deininger P L. Enhanced evolutionary PCR using oligonucleotides with inosine at the 3'-terminus. *Nucleic Acids Res*, 1991, 19: 5081
- 26 Shultz J W, Regier J C. Phylogenetic analysis of arthropods using two nuclear protein encoding genes supports a crustacean+hexapod clade. *Proc R Soc Lond B Biol Sci*, 2000, 267: 1011–1019
- 27 Candrian U, Furrer B, Hofelein C, et al. Use of inosine-containing oligonucleotide primers for enzymatic amplification of different alleles of the gene coding for heat-stable toxin type I of enterotoxigenic *Escherichia coli*. *Appl Environ Microbiol*, 1991, 57: 955–961
- 28 Christopherson C, Sninsky J, Kwok S. The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. *Nucleic Acids Res*, 1997, 25: 654–658
- 29 Folmer O, Black M, Hoeh W, et al. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Mol Marine Biol Biotechnol*, 1994, 3: 294–297
- 30 Song H, Buhay J E, Whiting M F, et al. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc Natl Acad Sci USA*, 2008, 105: 13486–13491
- 31 Kochhar H P S, Appa Rao K B C, Luciano A M, et al. *In vitro* production of cattle-water buffalo (*Bos taurus* – *Bubalus bubalis*) hybrid embryos. *Zygote*, 2002, 10: 155–162
- 32 Roth J, Myers P. *Bubalus bubalis*. Animal Diversity Web, 2004, [http://animaldiversity.ummz.umich.edu/site/accounts/information/Bubalus\\_bubalis.html](http://animaldiversity.ummz.umich.edu/site/accounts/information/Bubalus_bubalis.html)
- 33 Kerr K C R, Lijtmaer D A, Barreira A S, et al. Probing evolutionary patterns in neotropical birds through DNA Barcodes. *PLoS ONE*, 2009, 4: e4379

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Supporting Information

**Table S1** Collection localities and other information on the specimens

The supporting information is available online at [csb.scichina.com](http://csb.scichina.com) and [www.springerlink.com](http://www.springerlink.com). The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.