Animal Conservation

Animal Conservation. Print ISSN 1367-9430

A molecular diagnostic for identifying central African forest artiodactyls from faecal pellets

S. Ntie[1], A. R. Johnston[1], P. Mickala[2], A. E. Bowkett[3,4,5], B. Jansen van Vuuren[5], M. Colyn[6], P. Telfer[7,8], F. Maisels[7,9], O. Hymas[7], R. L. Rouyer[1], R. A. Wallace[1], K. LeBlanc[1], N. van Vliet[10], G. Sonet[11], E. Verheyen[11], D. Pires[12], E. J. Wickings[8], S. A. Lahm[13,14] & N. M. Anthony[1]

1 Department of Biological Sciences, University of New Orleans, New Orleans, LA, USA
2 Departement de Biologie, Université des Sciences et Techniques de Masuku, Franceville, Gabon
3 Whitley Wildlife Conservation Trust, Paignton Zoo, Paignton, UK
4 School of Biosciences, University of Exeter, Exeter, UK
5 Department of Botany and Zoology, University of Stellenbosch, Stellenbosch, Matieland, South Africa
6 Laboratory of Ethology, Evolution and Ecology, University of Rennes I, CNRS, UMR, Station Biologique de Paimpont, France
7 Wildlife Conservation Society, 2300 Southern Boulevard, New York, NY, USA
8 Centre International de Recherches Médicales de Franceville, Franceville, Gabon
9 School of Biological and Environmental Sciences, University of Stirling, Stirling, Scotland, UK
10 Center for International Forestry Research, Yaoundé, Cameroon
11 Royal Belgian Institute of Natural History, Brussels, Belgium
12 Life Sciences Core Curriculum, University of California, Los Angeles, CA, USA
13 Institut de Recherche en Ecologie Tropicale, Makokou, Gabon
14 Ecology and Environment Inc., Lancaster, NY, USA

Keywords
artiodactyl; duiker; control region; cytochrome *b*; species diagnostic; RFLP; barcoding; central Africa.

Correspondence
Nicola Anthony, Department of Biological Sciences, University of New Orleans, New Orleans, LA 70148, USA.
Email: nanthony@uno.edu

Received 3 January 2009; accepted 3 July 2009

doi:10.1111/j.1469-1795.2009.00303.x

Abstract

Small to medium-sized central African forest artiodactyls constitute a diverse yet heavily hunted group composed primarily of species within the genera *Cephalophus*, *Neotragus*, *Tragelaphus* and *Hyemoschus*. Of these genera, *Cephalophus* is the richest with as many as seven sympatric species known to occur in central African forests. However, differentiating species from their faeces or from tissue where the whole carcass is unavailable is very difficult. In order to develop a robust molecular diagnostic for species identification, a database of mitochondrial cytochrome *b* (553 bp) and control region (~675 bp) sequences was compiled from all forest *Cephalophus* species and other similarly sized, sympatric *Tragelaphus*, *Neotragus* and *Hyemoschus* species. Reference phylogenies from each marker were then used to recover the identity of sequences obtained from unknown faecal samples collected in the field. Results were then compared to determine which region best recovered species identity with the highest statistical support. Restriction fragment length polymorphisms (RFLPs) were also assessed as an alternative method for rapid species identification. Of the methods examined, tree-based analyses built on a geographically comprehensive database of control region sequences was the best means of reliably recovering species identity from central African duikers. However, three sister taxa appear indistinguishable (*Cephalophus callipygus*, *Cephalophus ogilbyi* and *Cephalophus weynsi*) and not all species were monophyletic. This lack of monophyly may be due to incomplete lineage sorting commonly observed in recently derived taxa, hybridization or the presence of nuclear translocated copies of mitochondrial DNA. The high level of intra-specific variation and lack of robust species-specific diagnostic sites made an RFLP-based approach to duiker species identification difficult to implement. The tree-based control region diagnostic presented here has many important applications including fine-scale mapping of species distributions, identification of confiscated tissue and environmental impact assessments.

Introduction

Forest artiodactyls are an important component of central African rainforest vertebrate communities and are widely distributed across the Congo basin (Feer, 1989*a*,*b*; Kingdon, 1997; Tutin, White & Mackanga-Missandzou, 1997). Many species are also an important source of protein and income for people living in the Congo basin (Lahm, 1993*a*; Wilkie &

Carpenter, 1999; Fa *et al*., 2002). For example, Wilkie & Carpenter (1999) reported that up to 95% of harvested bushmeat in Ituri forest within the Democratic Republic of the Congo (DRC) is made up of artiodactyls, namely forest duikers (genus *Cephalophus* spp.) and the bush pig *Potamochoerus porcus*. Similarly, Thibault & Blaney (2003) found in surveys conducted between 1997 and 1998 in Gamba market of Gabon that artiodactyls made up 62% of the total weight of species recorded. Among 254 animals killed by villagers in north-eastern Gabon, artiodactyls comprised 57.5%, of which duikers represented 86.3% (Lahm, 1993*a*). As a result, many African artiodactyls have become increasingly threatened by hunting pressure (Wilkie & Carpenter, 1999; Newing, 2001; Brashares *et al*., 2004; Laurance *et al*., 2006). Wildlife declines are also likely to be exacerbated by the rise in emerging infectious and epidemic diseases (Karesh *et al*., 1995; Food and Agricultural Organization, 1999; Lahm *et al*., 2007). Demographic projections suggest that if current levels of over-exploitation persist, several important forest artiodactyls will disappear within the next 100 years (Lahm, 1993*b*; Barnes, 2002).

The inability to correctly identify species and determine their proportional abundance in the wild is of real conservation concern, not only for species management but also in the regulation of illegal trade (Moore *et al*., 2003; Wasser *et al*., 2006; Fong *et al*., 2007). Within this context, improving our knowledge of species richness and abundance has become an increasingly important challenge in the development of long-term strategies for sustainable management of natural artiodactyl populations (Newing, 2001). To date, studies of tropical forest artiodactyl species have been traditionally based on line transect sampling of faecal piles (Walsh & White, 1999; White & Edwards, 2000; Ellis & Bernard, 2005), direct observations (Lahm, 1993*a*; Heydon & Bulloh, 1997; White & Edwards, 2000) and market sampling (Crooks, Ankudey & Milner-Gulland, 2005; Fa *et al*., 2006).

Despite their utility, each of these three approaches has proved to be problematic. Firstly, the identification of duiker species based on their faeces is unreliable (Bowkett *et al*., 2008; van Vliet *et al*., 2008). Faecal pellets from many sympatric artiodactyls are of comparable size and shape and cannot be used to reliably differentiate species in the field (Bowkett *et al*., 2008; van Vliet *et al*., 2008). Secondly, visual surveys remain difficult because most forest artiodactyls such as duikers are cryptic and elusive and may often require nocturnal surveys for best information (Feer, 1989*b*; Lahm, 1993*a*; Focardi, Isotti & Tinelli, 2002; Waltert *et al*., 2006; Croes *et al*., 2007). Lastly, practical identification of bushmeat samples may also be difficult when whole animal carcasses are not available. Census data based on market samples are likely to be incomplete because some species may not be sold due to cultural taboos or consumer preferences (Lahm, 1993*a*; van Vliet, 2008) and data on species geographic origin may be difficult to obtain (cf. Moore *et al*., 2003). Because numerous similar-sized artiodactyl species can occur in sympatry in central African tropical forests (Dubost, 1980, 1984; Kingdon, 1997; Newing, 2001; Lahm & Tezi, 2006), differentiating species becomes an even more challenging task.

Fortunately, polymerase chain reaction (PCR) methods now offer an alternative set of tools for obtaining genetic data from wild animal faeces or unidentified tissues, even when only degraded sources of DNA are available (Baker & Palumbi, 1994; Taberlet *et al*., 1996; Kohn *et al*., 1999; Dreher *et al*., 2007). Several molecular-based methods for biological species identification have been proposed. One approach is to build a reference phylogeny and then use the tree to assign samples to species-specific clades with high statistical support (e.g. Baker & Palumbi, 1994; Hammond *et al*., 2001; Ross *et al*., 2003; Lorenz *et al*., 2005; Baker *et al*., 2006). This only becomes a viable strategy if a substantive database of representative sequences from each target species has been obtained (Ekrem, Willassen & Stur, 2007) and the molecular marker used has sufficient resolution to reliably differentiate species with high bootstrap or posterior support. This approach is similar to a DNA 'barcoding' methodology where a standardized fragment of the mitochondrial genome is used for rapid species identification (Hebert *et al*., 2003). According to the DNA barcoding approach, species are identified by either one of the two following criteria: (1) reciprocal monophyly (Wiens & Penkroft, 2002) or less commonly (2) inter-specific genetic distances that are 10-fold or greater than the average intraspecific distance (Hebert *et al*., 2004). PCR amplified restriction fragment length polymorphisms (RFLP) have also been used as a simple yet cost-effective alternative to sequence-based approaches (Kohn *et al*., 1999; DeYoung & Honeycutt, 2005), especially where laboratory resources are limited (Chandiwana & Ornbjerg, 2003). However, the extent to which this method could be used to reliably diagnose geographically widespread and recently derived species remains poorly understood.

The goal of the present study is to use mitochondrial sequence data from a set of reference samples to develop a simple and efficient molecular diagnostic for the identification of unknown faecal or tissue samples from small to medium central African artiodactyl species. The four main aims of this study are therefore to: (1) build reference phylogenies based on mitochondrial cytochrome *b* gene and control region sequences using a set of tissue samples of known species identity; (2) compare the ability of these two markers to discriminate species and identify unknown faecal samples from different sites across central Africa, including a set of faecal samples diagnosed previously to species level using a 12S ribosomal RNA phylogeny (van Vliet *et al*., 2008); (3) evaluate the extent to which DNA barcoding criteria can be used to differentiate species; (4) assess the feasibility of using a RFLP-based approach to discriminate between target species in this study.

Methods

Sample collection and DNA extraction

Target species for this study comprise central African species in the genus *Cephalophus* (*C. monticola*,

C. callipygus, *C. dorsalis*, *C. silvicultor*, *C. nigrifrons*, *C. leucogaster* and *C. ogilbyi*) and other species in the sympatric genera *Neotragus* (*N. moschatus*), *Tragelaphus* (*T. spekei*, *T. scriptus*) and *Hyemoschus* (*H. aquaticus*). Because duikers constitute the most important group in the present study, all other species of *Cephalophus* whose range was outside central Africa were also included in phylogenetic analyses for taxonomic completeness. These taxa comprise: *C. rufilatus*, *C. natalensis*, *C. harveyi*, *C. spadix*, *C. adersi*, *C. maxwelli*, *C. niger*, *C. zebra*, *C. weynsi*, *C. rubidus, C. jentinki* and the closely related *Sylvicapra grimmia*. Reference tissue samples were donated from zoos, museums, scientific collections and bushmeat market surveys conducted in collaboration with the Wildlife Conservation Society in Gabon (Table 1). Samples obtained from Gabonese meat markets were collected from Franceville (FR), Okondja (OK) and Lamberéné (LAM). With the exception of the readily distinguishable blue duiker *C. monticola* and spiral horned antelope (*Tragelaphus* species), these bushmeat samples were accompanied by a photographic record for additional verification ($n = 22$). Similarly, teeth or tissue samples obtained from M. Colyn ($n = 25$) or from the San Diego Zoo ($n = 18$) were accompanied by either craniometric data (M. Colyn, unpubl. data) or species records, respectively. To assess the utility of the reference phylogeny as a species diagnostic, faecal DNA samples of unknown species identity were included from several sites across central Africa. These were: (1) Ossele (Oss), Ipassa (IP), Lopé National Park (Lope) and Monts de Cristal National Park (MCR) in Gabon; (2) Nouabalé-Ndoki National Park (ND) in the Republic of the Congo; (3) Salonga National Park (SA) in DRC; (4) Bioko island (BKO) in Equatorial Guinea; (5) Udzungwa Mountains (AB), in Tanzania. Details on the geographic location and donor(s) of these samples are listed in supporting information (Table S1).

DNA was extracted from tissue samples using a standard phenol–chloroform extraction method (Sambrook & Russell, 2001). Samples from museum skull scrapings and hide were extracted using the Geneclean® kit for ancient DNA (QBiogene, Carlsbad, CA, USA). For museum samples provided by M. Colyn, DNA was extracted from the pulp of a molar taken from the mandible of each skull. A portion of the root was removed by cutting through one or

Table 1 Samples sequenced in the present study

	Number of individuals			
Species	CR	Cytb	Total	Collectors
T. scriptus	4	5	6	Yoshan Moodley
T. spekei	5	1	5	Marc Colyn, Stevens Touladjan
H. aquaticus	7	2	7	Marc Colyn, Stevens Touladjan
S. grimmia	4	4	4	San Diego Zoo, Bettine Jansen van Vuuren
C. monticola	11	8	15	Debra Pires, Stevens Touladjan, Bettine Jansen van Vuuren, San Diego Zoo, American Museum of Natural History, Marc Colyn, Field Museum of Natural History, Genbank
C. maxwelli	3	3	4	Genbank, Field Museum of Natural History, San Diego Zoo
C. adersi	2	3	3	Bettine Jansen van Vuuren, Genbank
C. natalensis	3	3	4	Bettine Jansen van Vuuren, Genbank, Field Museum of Natural History,
C. harveyi	8	3	9	Bettine Jansen van Vuuren, Andrew Bowkett, Genbank,
C. rufilatus	7	4	8	San Diego Zoo, Bettine Jansen van Vuuren, Genbank
C. nigrifrons	5	4	6	Debra Pires, Bettine Jansen van Vuuren, Genbank
C. zebra	2	2	3	San Diego Zoo, Marc Colyn, Genbank
C. leucogaster	5	3	6	Bettine Jansen van Vuuren, Debra Pires, Stevens Touladjan, Genbank
C. callipygus	9	7	12	Bettine Jansen van Vuuren, Stevens Touladjan, Deb Pires, Genbank
C. niger	3	3	5	Andrew Bowkett, San Diego Zoo, Bettine Jansen van Vuuren, Genbank, Field Museum of Natural History,
C. spadix	8	3	9	Bettine Jansen van Vuuren, Andrew Bowkett, Genbank
C. silvicultor	7	5	9	Debra Pires, San Diego Zoo, Bettine Jansen van Vuuren, Marc Colyn, Genbank
C. ogilbyi	4	4	5	Marc Colyn, Stevens Touladjan, Genbank
C. dorsalis	21	8	26	Stevens Touladjan, Debra Pires, San Diego Zoo, American Museum of Natural History, Field Museum of Natural History, Marc Colyn, Genbank,
C. walteri	2	0	2	Marc Colyn
C. weynsi	1	2	2	Bettine Jansen van Vuuren, Genbank
C. jentinki	0	1	1	Genbank
C. rubidus	0	1	1	Genbank
Faeces of provisional identity	9	11	11	Natalie van Vliet *et al.* (2008)
Faeces of unknown identity	15	13	27	Bryan Curran, Andrew Bowkett, Emma Stokes, Fiona Maisels, Patrick Mickala, R. Aba Nzenme, J. Larry Dew, Anne Johnston, Stephan Ntie

CR, control region; Cytb, cytochrome *b* gene.

more cusps with a sterilized drill and then incubated overnight at 65 °C in lysis buffer provided in QIAamp DNA Blood Minikit (Qiagen, Valencia, CA, USA). DNA extraction was then carried out according to the manufacturer's instructions. Faecal DNA extractions were carried out in a room designated for non-invasive DNA procedures at the Centre International de Recherches Médicales de Franceville (CIRMF), Gabon, using the QIAamp DNA Stool Minikit (Qiagen). A blank was included in each extraction series to control for DNA contamination.

PCR amplification

Primers CytbF1 and CytbR2 were designed from available GenBank sequence data and were used to amplify a 553 bp fragment of the mitochondrial cytochrome *b* gene (Table 2). Alternatively, a 1140 bp fragment encompassing the entire cytochrome *b* gene was amplified using primers L14724 and H15915 (Pääbo & Wilson, 1988). For the mitochondrial control region, we initially amplified the entire region using primers CRF1 and CRR1 located in the flanking tRNA genes. These primers were based on available GenBank data from *Cephalophus* (AJ235317) and *Neotragus* (AJ235323) species. As the initial *Cephalophus* and *Tragelaphus* sequence data amplified using this primer pair indicated that the left hand domain was more variable, we focused subsequent analyses on ~675 bp fragment encompassing this domain by using an internal primer CRR3 in combination with CRF1 (Table 2). Alternatively, a slightly smaller, overlapping (~600 bp) fragment was amplified using primers N777 (modified from Hoelzel, Hancock & Dover, 1991) and H16498 (Shields & Kocher, 1991). In cases where sample DNA was degraded, internal primers were designed to amplify the corresponding region in smaller, overlapping fragments. For the cytochrome *b* gene, the internal reverse primer CytbR1 and internal forward primer CytbF2 were used in combination with CytbF1 and CytbR2 primers, respectively. Similarly, internal control region primers CRR5A and CRF6A were used in combination with primers CRF1 and CRR3, respectively (Table 2).

PCR reactions were carried out in a 50 μL reaction volume containing 1 × enzyme buffer (200 mM Tris pH 8.4, 500 mM KCl), 1.5–3.0 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 μM of each primer, 1.25 U of *Taq* polymerase (Invitrogen, Carlsbad, CA, USA) and 1–2 μL of template DNA. DNA amplifications were carried out using the GeneAmp® 9700 thermocycler (Applied Biosystems, Foster City, CA, USA) or I-cycler (Biorad, Hercules, CA, USA) with an initial denaturation time of 94 °C for 4 min, followed by 35 cycles of initial denaturation at 94 °C for 30 s, annealing at 50–55 °C for 30 s, and an extension step at 72 °C for 30 s. The programme ended with a final extension step of 72 °C for 5–10 min. A negative control containing water instead of template DNA was included in each amplification reaction. PCR products were purified using the Geneclean® Turbokit (QBiogene) according to the manufacturer's instructions. Sequencing was carried out on an automated DNA sequencer (ABI 3100) using the BigDye Terminator Cycle Sequencing Kit v1.1 (Applied Biosystems).

Phylogenetic analysis

Sequence data were edited using the program SEQUENCHER v 4.1.1 (Gene Codes Corporation, Ann Arbor, MI, USA) and aligned before phylogenetic analysis using Clustal X v2.06 (Thompson *et al.*, 1997). As control region sequence alignment can be problematic, we also aligned this dataset using a higher performance procedure implemented in MUSCLE (Edgar, 2004). Unlike the cytochrome *b* dataset, control region sequences from *H. aquaticus* and *Tragelaphus* species were omitted due to their high divergence from *Cephalophus* species in the present study although exploratory analyses were also conducted with *Tragelaphus* sequences included. A highly variable section of the Clustal X control region alignment containing a large insert in some *C. monticola* sequences was also deleted before phylogenetic analysis to determine whether this affected the overall tree topology

Table 2 Primer pairs and reaction conditions for cytochrome *b* and control region amplification

Primer	Target region	Sequence	Primer pair	$[Mg^{2+}]$ (mM)	Annealing temperature (°C)
Cytb F1	Cytochrome *b*	5′-TACATACACGCAAACGGAGC-3′	F1/R1	1.5	52
Cytb R1	Cytochrome *b*	5′-TGGAAGGCGAAGAATCGG-3′			
Cytb F2	Cytochrome *b*	5′-TCTGAGGGGGCTTTTCAGTAG-3′	F2/R2	2.5	50
Cytb R2	Cytochrome *b*	5′-TGTGTTGAGTGGGTTTGC-3′	F1/R2	1.5	51
L14724	Cytochrome *b*	5′-CGAAGCTTGATATGAAAAACCATCGTTG-3′			
H15915	Cytochrome *b*	5′-CGAGATCTGAAAAACCATCGTTG-3′	L14724/H15915		50
CR F1	Control region	5′-CTCCCTAAGGCTCAAGGAAGC-3′	F1/R3	3.0	56
CR R1	Control region	5′-CATCTAGGCATTTTCAGTGCC-3′	F1/R1	1.5	55
CR R3	Control region	5′-CCTGAAGRAAGAACCAGATGTC-3′			
CR R5A	Control region	5′-CATTAATCCTTGTTGTACTTGC-3′	F1/R5A	2.0	53
CR F6A	Control region	5′-GTTATACAGACATACTATGTATATAG-3′	F6A/R3	1.5	53
N777	Control region	5′-TACACTGGTCTTGTAAACC-3′	N777/H16498	1.5	48
H16498	Control region	5′-CCTGAAGTAGGAACCAGATG-3′			

and recovery of species identity. Lastly, in order to test for the potential presence of non-functional nuclear translocated copies of mitochondrial DNA (Numts), cytochrome *b* gene sequence data were translated using the program MEGA v3 (Kumar, Tamura & Nei, 2004) and examined for evidence of frameshifts or stop codons.

Phylogenetic analyses of both cytochrome *b* ($n = 106$) and control region ($n = 138$) datasets was carried out using the neighbor-joining (NJ) and maximum parsimony (MP) methods implemented in PAUP 4.0b10 (Swofford, 2000), or the Bayesian method used in MrBayes (Huelsenbeck & Ronquist, 2001). In NJ analyses, a Kimura-2-parameter model of nucleotide substitution was adopted, as is recommended for species-level DNA barcoding analyses (Hebert *et al.*, 2003). For MP analyses, a starting tree was obtained using the stepwise addition option and heuristic searches were conducted using the tree–bisection–reconnection (TBR) algorithm. All character changes were considered unordered and unweighted. Bayesian analyses were carried out using the Monte Carlo Markov Chain (MCMC) method implemented in MrBayes and a general time-reversible model that allowed for among site rate variation and invariant sites. Prior probabilities for model parameters were left at their default settings. In order to ensure that the MCMC chain had not been trapped in local optima (Leaché & Reeder, 2002), output was compared from two separate analyses, each made up of three heated chains and a cold chain. The proportion of samples to be discarded as 'burn in' was assessed by looking at the output from the *sump* command in MrBayes and by examining the MCMC trace files using the program TRACER (Rambaut & Drummond, 2007). In each case, runs were only accepted if the effective sample size (ESS) was >150 for all model parameters. Convergence was assessed by verifying whether different runs attained the same stationary distribution and average log-likelihood values and attained a standard deviation of split frequencies of <0.01. Chains were run for 10 000 000–60 000 000 iterations and trees were sampled every 10 000 generations. Support for a specific node was accepted if the relevant bootstrap value was $\geq 75\%$ and posterior probabilities were ≥ 0.95.

Assessment of DNA barcoding criteria

DNA barcoding criteria were also applied to the cytochrome *b* gene dataset in order to assess whether this region could be used to identify species based on the criterion of either reciprocal monophyly or the 10-fold genetic distance rule. Although cytochrome *b* is not used for DNA barcoding studies, like the cytochrome *c* oxidase I gene, it encodes a functional, polymorphic protein and has been used to resolve species level phylogenies in the genus *Cephalophus* (Jansen van Vuuren & Robinson, 2001). Pair-wise Kimura-2-parameter genetic distances were calculated using the program MEGA v3 (Kumar *et al.*, 2004). The resulting data matrix was then used to construct a NJ tree and assess the extent to which *Cephalophus* taxa conform to the 10-fold genetic distance rule.

Cytochrome *b* and control region RFLP analysis

RFLP analyses were restricted to species that are known to occur sympatrically in central African rainforests. These species comprise: *C. callipygus*, *C. dorsalis*, *C. leucogaster*, *C. monticola*, *C. nigrifrons*, *C. ogilbyi*, *C. silvicultor*, *T. scriptus*, *T. spekei* and *H. aquaticus*. Potential species-specific restriction enzymes were identified by mapping candidate restriction enzymes to cytochrome *b* and control region sequence alignments using the program DSGene (Accelrys, San Diego, CA, USA). Enzymes were only selected for RFLP analysis if they contained one or two cut sites unique to all reference individuals within a target species, possessed recognition sites of 5–6 bp in length and were commercially available through New England Biolabs (Ipswich, MA, USA) or Invitrogen. The reference database used to screen for RFLPs was identical to that used to build phylogenetic trees for faecal sample identification.

Results

Phylogenetic analysis

With respect to the cytochrome *b* data, Bayesian analysis provided strong support for *C. dorsalis*, *C. zebra*, *C. niger*, *C. adersi*, *S. grimmia*, *N. moschatus* and *H. aquaticus* species clades (Fig. 1). In contrast, support was weak or absent for most of the central African species that constitute the target of this study and several sister taxa could not be discriminated from one another. Specifically (1) *C. nigrifrons* could not be distinguished from *C. rufilatus*, *C. harveyi* and *C. natalensis*; (2) *C. ogilbyi* could not be distinguished from either *C. callipygus* or *C. weynsi*; (3) *C. monticola* and *C. maxwelli* were unresolved. Furthermore, *C. leucogaster* appeared paraphyletic with respect to a multi-species clade consisting of *C. nigrifrons*, *C. rufilatus*, *C. natalensis* and *C. harveyi*. Similarly, *T. scriptus* is paraphyletic with respect to its sister taxon *T. spekei*, as observed previously (Moodley *et al.*, 2009). Owing to the lack of resolution, the identity of almost all faecal DNA samples could not be reliably recovered, with the exception of faecal samples from *C. dorsalis*. Both cytochrome *b* NJ (supporting information Fig. S1) and MP (supporting information Fig. S2) analyses also failed to resolve many central African species clades and in the case of MP resolved even fewer species nodes than either of the other two methods.

In contrast, Bayesian analyses of the control region dataset using either alignment procedure were not only able to successfully resolve target taxa but did so with high posterior support (Fig. 2). The only exceptions to this observation are samples from the sister taxa *C. callipygus* and *C. ogilbyi*, both of which fell in two different clades (A, B). The geographically localized species *C. weynsi* also fell within clade A but nevertheless appeared to constitute a distinct haplotype. Unlike the cytochrome *b* phylogeny, *C. nigrifrons* could be distinguished from *C. rufilatus* with high posterior support. However, neither species constituted

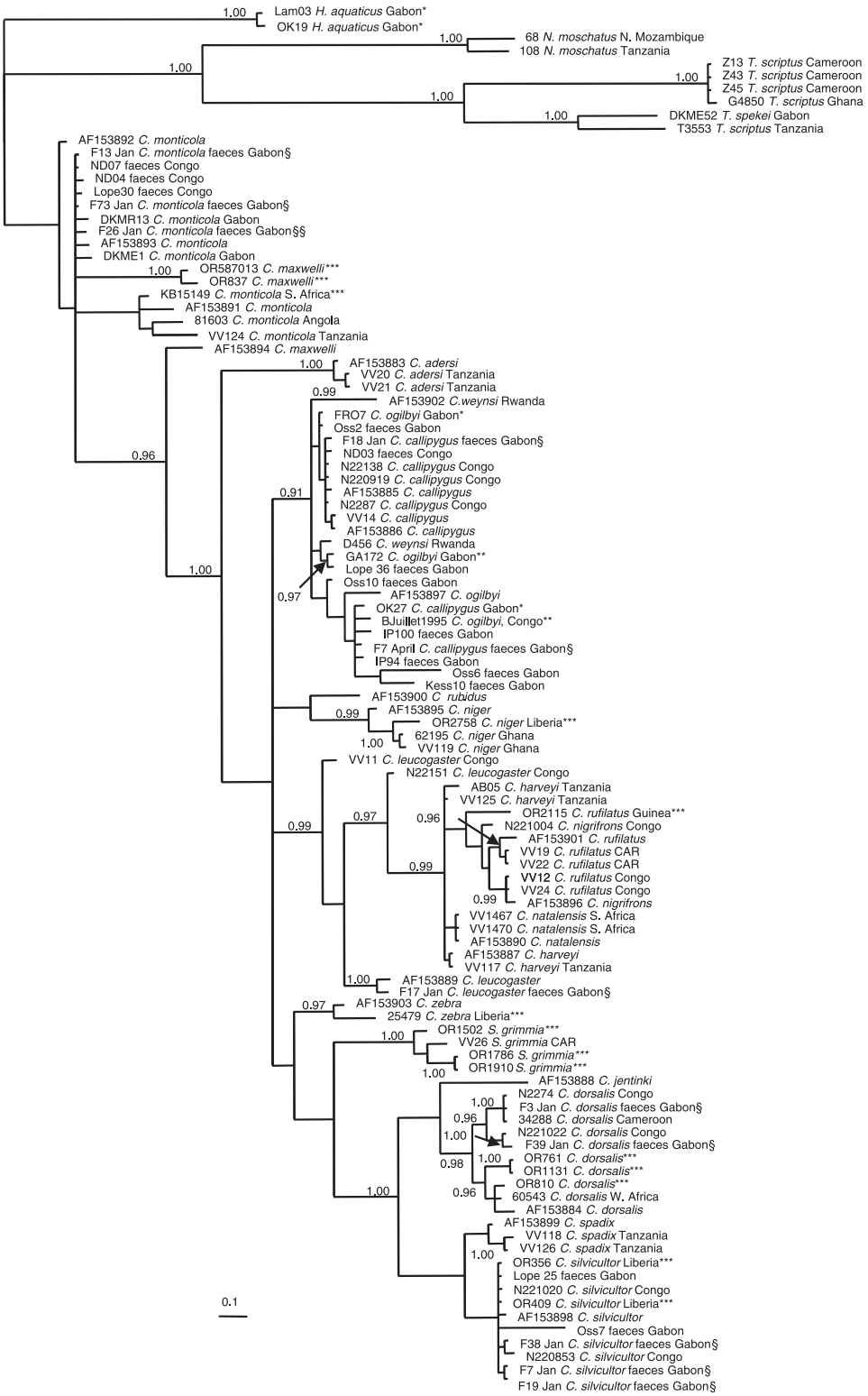

Figure 1 Cytochrome *b* phylogeny based on Bayesian analysis and rooted with *Hyemoschus aquaticus*. Sequences were aligned with Clustal X 2.06. Posterior probability values are indicated at the appropriate node. Specimens that have been verified by either: *photographing the whole animal; **craniometric analysis; ***currently held at the Centre for Research on Endangered Species at San Diego Zoo are marked appropriately. §Samples obtained from van Vliet *et al.* (2008).

a monophyletic association, but instead appeared to be made up of two distinct clades (A, B). Other species that did not constitute monophyletic associations comprised: (1) *C. monticola* that appeared paraphyletic to *C. maxwelli*; (2) *C. harveyi* that appeared paraphyletic to *C. natalensis*. Regardless of the lack of monophyly of some species, the control region phylogeny was always able to recover the species identity of the faecal samples examined in this study. This finding was observed regardless of whether the *Tragelaphus* species sequences were included (supporting information Fig. S3) or excluded (Fig. 2) as outgroups in the control region phylogeny. Similarly, deletion of a section of the control region alignment encompassing the *C. monticola* clade B insert or use of a different alignment method had no effect on the high levels of posterior support observed for individual species clades in Bayesian analysis. Both NJ and MP analyses of the control region dataset also resolved all species nodes with high bootstrap support (supporting information Figs S4 and S5). However, MP analysis of the MUSCLE alignment only gave weak bootstrap support ($<75\%$) for *C. spadix* and the *C. monticola* clade A. Lastly, control region analyses presented here confirmed the provisional species identification of unknown fecal samples made by van Vliet *et al.* (2008).

We found no evidence of multiple peaks in the chromatogram of either marker or frameshifts and stop codons in the cytochrome *b* gene, which would indicate the presence of Numts. However, the *C. callipygus*/*C. ogilbyi* clade B (Fig. 2) bears some of the possible hallmarks of a Numt group. Firstly, there is little to no variation within this clade, consistent with Zischler *et al.*'s hypothesis (1995) that once a mitochondrial haplotype is translocated it becomes 'fossilized' due to the much slower mutation rate in the nuclear genome. Secondly, the pronounced differentiation between the control region clades A and B is not observed in the cytochrome *b* phylogeny (Fig. 1). Lastly, the *C. callipygus*/*C. ogilbyi*/*C. weynsi* clade A exhibits a comparable level of intra-specific diversity to the corresponding species clade in the cytochrome *b* phylogeny suggesting that both are mitochondrial in origin.

Several samples were removed from the control region phylogenetic analysis because their location in the phylogeny most likely resulted from specimen misidentification, mislabelling or contamination from another species source. These comprise: the GenBank sequence from *C. leucogaster* AJ235317, a sample of *H. aquaticus* (FR16) that did not match its photographic record and a sample of *C. niger* (VV131) that falls within the *C. maxwelli* clade. Museum samples are also highly susceptible to contamination and several cases of smaller overlapping PCR fragments that did not match one another (AMNH 140902, AMNH 60543, VV *C. ogilbyi*) or appeared contaminated by another, unrelated species (T14 *C. zebra*) were detected. We also found one instance where duplicate extractions from the same animal (0108-1DOR, R16492) did not match one another. One faecal sample from van Vliet *et al.* (2008) was most likely mislabelled because it was originally diagnosed as *C. nigrifrons* in the 12S ribosomal RNA phylogeny of van Vliet *et al.* (2008) but was indistinguishable from a *C. callipygus* sample (F18 Jan) included in the present analysis. Two geographically restricted species (*C. rubidus*, *C. jentinki*) also failed to amplify, possibly due to the degraded condition of these samples. Lastly, there was one sequence anomaly in the cytochrome *b* database where two different haplotypes were obtained from the same sample of *C. weynsi* (Fig. 1). The best explanation is that one of these sequences is a contaminant or a numt although it is impossible to differentiate these hypotheses at present. Regardless, both cluster within the same clade so that their differentiation does not affect our conclusions. Despite these few instances of sample mislabelling and contamination in the control region database ($10/138 = 7.2\%$), it is important to note that in all other cases, reference samples in the control region phylogeny always fell into the correct species clade(s) and with high statistical support (Fig. 2).

Evaluation of DNA barcoding criteria

The NJ tree constructed from Kimura-2-parameter distances for the cytochrome *b* dataset (supporting information Fig. S1) recapitulate observations made from both MP and Bayesian analyses in that some but not all species constitute monophyletic associations. An examination of the between versus within species Kimura-2-parameter distances also illustrates that few species comparisons satisfy the 10-fold genetic distance rule, even between species within different genera (Table 3). This problem may be particularly acute in species with high intra-specific genetic variation (e.g. *C. dorsalis*), where one or more paraphyletic lineages are present (e.g. *T. scriptus*) or where Numts are inadvertently amplified (Song *et al.*, 2008).

Cytochrome *b* and control region PCR-RFLP analysis

A cytochrome *b* gene RFLP diagnostic was developed (supporting information Table S2) and a flow chart was designed from the combination of enzymes that together could be used for species identification (supporting information Fig. S6). However, enzyme digests yielded inconsistent results with tissue samples or unknown samples suggesting that within for species diversity was too high to be able to successfully implement this assay. Similar results were obtained with the control region RFLP analyses in that initial analysis of a subset of sequences from all central African target artiodactyls showed several candidate restriction enzyme sites that could be used to distinguish species (supporting information Table S3). However, additional polymorphisms at these enzyme restriction sites emerged as more sequences were added to the database making this assay difficult to implement. Despite the failure to develop a reliable RFLP diagnostic, a $\sim$275 bp deletion within the control region of *H. aquaticus* samples could be used to rapidly identify this species without sequence data. Similarly, one of the *C. monticola* clades (B) possesses an insertion of $\sim$60 bp in length that can be recognized on an agarose gel. Aside from these two cases, the identification of faecal samples in this study could only be made through sequencing and phylogenetic analysis.

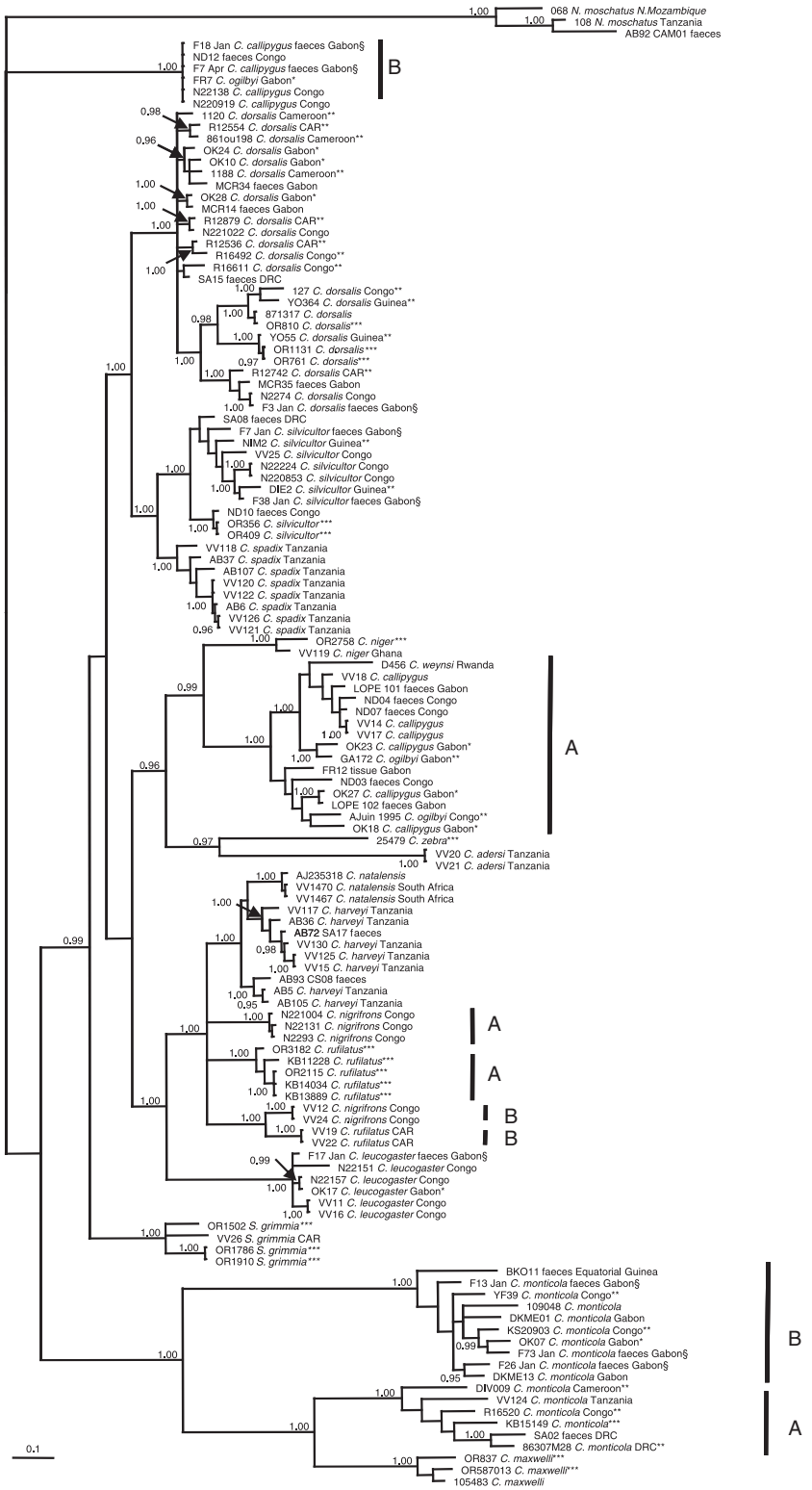

Figure 2 Control region phylogeny based on Bayesian analysis and rooted with *Neotragus moschatus*. Sequences were aligned with Clustal. Posterior probability values are indicated at the appropriate node. Specimens that have been verified by either: *Photographing the whole animal; **craniometric analysis; ***currently held at the Centre for Research on Endangered Species at San Diego Zoo are marked appropriately. Note that letters (A, B) denote two different clades within *Cephalophus callipygus*, *Cephalophus nigrifrons*, *Cephalophus rufilatus* and *Cephalophus monticola* that differ in their restriction fragment length polymorphism banding profiles. §Samples obtained from van Vliet *et al.* (2008).

Discussion

With the exception of the sister species *C. callipygus*, *C. ogilbyi* and *C. weynsi*, phylogenetic analysis of control region sequences was highly successful in recovering species identities of faecal samples. The only other study to date that has used a phylogenetic approach to identify faeces from central African duikers is that of van Vliet (2008). This

Table 3 Pair-wise cytochrome *b* gene Kimura 2-parameter corrected distances for all taxa included in the present study

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.
1. *C. silvicultor*	**0.006**																						
2. *C. spadix*	0.018	**0.005**																					
3. *C. dorsalis*	0.062	0.057	**0.017**																				
4. *C. jentinki*	0.056	0.058	0.042	**0.000**[a]																			
5. *C. callipygus*	0.081	0.093	0.077	0.066	**0.007**																		
6. *C. ogilbyi*	0.083	0.096	0.076	0.066	0.014	**0.019**																	
7. *C. weynsi*	0.071	0.084	0.067	0.056	0.021	0.022	**0.026**																
8. *C. rubidus*	0.070	0.078	0.074	0.053	0.038	0.041	0.042	**0.000**[a]															
9. *C. rufilatus*	0.088	0.094	0.076	0.072	0.059	0.060	0.058	0.063	**0.013**														
10. *C. nigrifrons*	0.080	0.087	0.067	0.065	0.053	0.054	0.053	0.055	0.015	**0.008**													
11. *C. natalensis*	0.077	0.090	0.080	0.062	0.053	0.053	0.053	0.051	0.023	0.022	**0.000**												
12. *C. harveyi*	0.081	0.094	0.083	0.065	0.059	0.060	0.060	0.057	0.018	0.021	0.006	**0.001**											
13. *C. leucogaster*	0.070	0.079	0.069	0.058	0.057	0.059	0.051	0.054	0.037	0.036	0.024	0.030	**0.015**										
14. *S. grimmia*	0.071	0.073	0.058	0.053	0.075	0.077	0.067	0.062	0.072	0.066	0.071	0.075	0.065	**0.016**									
15. *C. zebra*	0.050	0.062	0.057	0.055	0.050	0.048	0.045	0.049	0.060	0.051	0.049	0.053	0.050	0.061	**0.018**								
16. *C. niger*	0.087	0.089	0.083	0.081	0.062	0.064	0.068	0.042	0.072	0.063	0.068	0.074	0.068	0.083	0.061	**0.017**							
17. *C. adersi*	0.065	0.076	0.087	0.086	0.082	0.084	0.081	0.061	0.086	0.078	0.064	0.068	0.075	0.081	0.061	0.078	**0.002**						
18. *C. maxwelli*	0.074	0.086	0.084	0.082	0.094	0.096	0.082	0.087	0.088	0.079	0.082	0.087	0.078	0.093	0.069	0.087	0.080	**0.048**					
19. *C. monticola*	0.081	0.089	0.093	0.086	0.090	0.091	0.084	0.086	0.093	0.083	0.089	0.093	0.083	0.090	0.073	0.083	0.086	0.054	**0.033**				
20. *H. aquaticus*	0.144	0.160	0.144	0.139	0.139	0.141	0.137	0.135	0.134	0.128	0.135	0.139	0.133	0.137	0.118	0.135	0.143	0.123	0.112	**0.005**			
21. *N. moschatus*	0.140	0.147	0.158	0.136	0.151	0.149	0.145	0.132	0.157	0.151	0.145	0.149	0.142	0.151	0.126	0.144	0.153	0.148	0.132	0.165	**0.026**		
22. *T. scriptus*	0.161	0.167	0.161	0.148	0.152	0.154	0.155	0.149	0.154	0.147	0.144	0.146	0.149	0.161	0.137	0.152	0.172	0.162	0.156	0.168	0.159	**0.042**	
23. *T. spekei*	0.140	0.148	0.159	0.140	0.150	0.155	0.153	0.134	0.152	0.148	0.137	0.144	0.140	0.141	0.131	0.165	0.152	0.148	0.147	0.148	0.182	0.102	**0.000**[a]

Within species distances are on the diagonal and in bold.

[a]An estimate based on only one sample.

earlier study used a mitochondrial 12S ribosomal RNA phylogeny to recover species identity from unknown faecal samples collected from a single site in Gabon. Although the identifications made were consistent across both studies, it is important to note that the 12S dataset lacked a broad reference database and failed to adequately resolve species such as *C. dorsalis*, *C. silvicultor* and *C. callipygus* with acceptably high levels of Bayesian posterior support (>0.95). A similar lack of species resolution was also evident in our cytochrome *b* phylogeny and reflects the unresolved phylogenetic relationships evident between some species in earlier work by Jansen van Vuuren & Robinson (2001).

The finding that several species are not monophyletic underscores the importance of drawing samples from a broad geographic range as possible in order to gauge the full extent of variability within a given species (Ekrem *et al.*, 2007). This lack of monophyly within several species may reflect the likely recent origin of many taxa that formed the target of this study or result from hybridization between sister taxa. Whereas the dwarf duiker clade (*C. monticola* and *C. maxwelli*) is estimated to have diverged ~5 million-years ago from other *Cephalophus* species, the remaining species within this genus are believed to have arisen relatively recently (Vrba, 1995; Jansen van Vuuren & Robinson, 2001). Numts (Bensasson *et al.*, 2001) may also explain the lack of monophyly in some species such as *C. callipygus* and its sister taxon *C. ogilbyi*. Although frameshifts and/or stop codons can be used to identify candidate Numts from coding regions of the mitochondrial genome, these criteria do not apply to non-coding regions such as the control region, making Numts from this region especially problematic to detect (e.g. Anthony *et al.*, 2007).

It is important to emphasize that neither the cytochrome *b* nor control region phylogenies can be used to adequately resolve phylogenetic relationships within the genus *Cephalophus*. In the case of the cytochrome *b* dataset, this might be due to the inadequate phylogenetic signal inherent to the short fragment used in the present study. In the case of the control region, it is more likely due to the difficulties of aligning divergent sequences between more distantly related genera and the homoplasy that may result from this. For example, when *Tragelaphus* was used as an outgroup, the *C. monticola*/*C. maxwelli* clade was not basal to the rest of the species within the genus *Cephalophus* despite its supposed earlier origin (Jansen van Vuuren & Robinson, 2001). Future work will accordingly assess phylogenetic relationships between *Cephalophus* species using a suite of nuclear introns selected to resolve phylogenetic relationships between closely related bovid species (Willows-Munro, Robinson & Matthee, 2005). These data may also prove useful in determining whether the non-monophyletic associations of some species in the control region phylogeny are due to Numts. If the same branching patterns and species-specific clades are evident in both datasets we can exclude the possibility of Numt contamination. However, if nuclear phylogenies fail to recapitulate the two clades evident in some species then we can conclude that either there is insufficient information to adequately resolve these clades, that hybridization has occurred or that one mitochondrial clade is of nuclear origin.

Findings from this study also raise the question as to whether a DNA barcoding approach could be successfully applied to species within the genus *Cephalophus*, given that this group is geographically widespread and recently derived (Moritz & Cicero, 2004; Hickerson, Meyer & Moritz, 2006). Pair-wise comparisons of Kimura-2-parameter distances between taxa indicate that almost all *Cephalophus* species do not satisfy the 10-fold genetic distance rule described by Hebert *et al.* (2004). This is due to both the high intraspecific diversity present in many species and their recent divergence from one another. The few cases where the 10-fold criterion is satisfied are likely to be biased by the poor geographic representation of samples within a given taxon (e.g. *C. rubidus*, $n = 1$). Future work should therefore evaluate how well cytochrome *c* oxidase I gene barcodes can differentiate central African artiodactyls and under what circumstances DNA barcoding can be used to identify species that have undergone recent, rapid radiations such as those within the genus *Cephalophus*.

Although we did not explicitly test many of the alternative molecular methods of species diagnosis, few if any represent viable alternatives to the control region phylogeny used in the present study. For example, although single-strand conformation polymorphism analysis (Orita *et al.*, 1989) has proved to be an effective, low-cost method for the identification of previously diagnosed haplotypes (Travis & Keim, 1995), this method quickly becomes unmanageable in cases of high haplotype diversity because reference samples always need to be run on the same gel. Moreover, bands migrating at the same position may not necessarily be homologous. It has also been suggested that microsatellite loci be used to identify species either through the presence/absence of species-specific alleles (Pilot *et al.*, 2007) or through assignment methods (Vàzquez-Domínguez *et al.*, 2001). However, the high cost of multiplex development and need to replicate genotypes many times when using non-invasive sources of DNA (Taberlet *et al.*, 1996; Taberlet & Luikart, 1999) makes this method prohibitively expensive and time consuming. Bowkett *et al.* (2008) successfully used a multivariate analysis of pair-wise genetic distances to resolve east African duiker species. However, this method quickly becomes intractable in datasets with large numbers of closely related taxa, as is the case here.

There are cases where a RFLP approach has been used successfully to diagnose species. For example, Moore *et al.* (2003) used a RFLP approach based on the mitochondrial cytochrome *b* gene to differentiate sea turtle species. Although a species-specific banding profile was successfully identified for all seven target sea turtle species, geographic sampling within the taxa examined was limited. In contrast, findings from the present study demonstrate that the RFLP approach built on either the cytochrome *b* gene or control region cannot be used to identify recently derived species such as those within the genus *Cephalophus*. However, this problem was only evident after substantially increasing sample representation within our reference database. The RFLP diagnostic likely failed as a result of the high amounts

of intra-specific genetic variation within species and the absence of available species-specific restriction sites, due in part to the recent and rapid radiation of this species complex (Vrba, 1995). However, an RFLP-based approach may work well for more distantly related taxa (Bidlack *et al*., 2007) or for differentiating species within specific geographic locales (Zapata *et al*., 2007).

One potential drawback to the present work is that some samples in the present reference database cannot be verified with voucher specimens. When questions over misidentification arise, as is the case for a few of our samples, it is practically impossible to cross-verify the identity of the specimen without having photo-verification or access to the voucher. It is worth noting, however, that samples whose identity had been supported by additional information (i.e. photo-verification at the time of collection, craniometric analysis or zoo records), always fell into the appropriate species clade. Nevertheless, there are several species lineages where external verification is lacking, drawing attention to the need for voucher museum specimens in the future.

In summary, the control region phylogeny presented here is the first tree-based method to be able to successfully differentiate sympatric central African artiodactyl species with high statistical support using a broad geographic reference database. Based on these and earlier findings (Focardi *et al*., 2002; Waltert *et al*., 2006; Croes *et al*., 2007; Bowkett *et al*., 2008; van Vliet *et al*., 2008), we therefore recommend using this marker to accurately recover species identity. However, while the control region phylogeny can be used to identify unknown samples to species level, this approach should not be used to evaluate taxonomic relationships within the genus *Cephalophus*. Additional data is needed to further examine their systematic relationships, especially since the present study indicates considerable ambiguity in the recognition of several species (e.g. *C. ogilbyi*, *C. weynsi*).

Although we used the reference diagnostic here to identify faecal samples, this approach could readily be applied to tissue fragments and other wildlife remains. This diagnostic also has numerous practical applications to conservation and management of African artiodactyl communities including: (1) mapping habitat preferences and species distributions; (2) identifying bushmeat samples from urban and regional markets; (3) estimating species abundance when combined with individual multi-locus genotype data; (4) predicting species responses to habitat modification through logging, fragmentation and intensive hunting; (5) determining species presence and relative abundance in environmental impact assessments (EIAs) of development projects and in associated in wildlife monitoring programmes. Owing to the rapid worldwide expansion of industrial and commercial development, many wildlife populations are increasingly contained within large concessions and thus need to be surveyed and managed as if they were in protected areas (Lahm & Tezi, 2006). The molecular diagnostic tool developed here may then prove particularly valuable for rapid species assessments and EIAs where rare, elusive and/or sparsely distributed species may otherwise be overlooked.

Acknowledgements

The authors wish to thank all those contributors to the sample reference database in Table 1 especially Oliver Ryder (Zoological Society of San Diego) and Mitchell Eaton (University of Colorado). We also wish to thank the Conseil National des Parcs Nationaux (CNPN), the Ministère de l'Environnement et de la Protection de la Nature, the Ministère de la Recherche Scientifique et du Développement Technologique, the Centre National de la Recherche Scientifique et Technologique (CENAREST) within the Ministère de l'Enseignement Supérieur et de l'Innovation Technologique, and the Département de la Faune et de la Chasse within the Ministère de l'Économie Forestière for permission to conduct this research in Gabon. We would like to also thank the Institut de Recherche en Ecologie Tropicale (IRET), Gabon, and the Centre International de Recherches Médicales de Franceville (CIRMF), Gabon, for logistical support and field assistance and the conservators Mr Ngouwou of Lopé National Park and Mr Allogho of the Monts de Cristal National Park for permission to collect faecal samples from both national parks in Gabon. We also wish to thank the Ministry of Forest Economy and the Department of Wildlife and Protected Areas in the Republic of the Congo, especially M. Onday and B. Djoni, and the WCS Nouabalé-Ndoki Project for support in sample collection. We also thank the Institut Congolais pour la Conservation de la Nature (ICCN), WCS and WWF DRC, and the MIKE-CITES programme for help in the Salonga National Park, DRC. Similarly, we would also like to thank the director of the Ministerio de Agricultura y Bosques Intituto Nacional de Desarrollo Foresta y Gestion de Sistema de Areas Protegidas (INDEFOR) of Equatorial Guinea for permission to collect samples and for their logistical support. A.E.B. would also like to thank Jamie Stevens and Amy Plowman for supervision, support and advice. G.S. would like to thank the Joint Experimental Molecular Unit (JEMU) founded by the Belgian Science Policy. This study was funded by NSF grant DEB 0516425 and the Coopération Française, Ministère des Affaires Etrangères, France (#FSP Titre VI Projet 2002 0046 00 FORINFO).

References

Anthony, N.M., Clifford, S.L., Bawe-Johnson, M., Abernethy, K.A., Bruford, M.W. & Wickings, E.J. (2007). Distinguishing gorilla mitochondrial sequences from nuclear integrations and PCR recombinants: guidelines for their diagnosis in complex sequence databases. *Mol. Phylogenet. Evol.* **43**, 553–566.

Baker, C.S. & Palumbi, S.R. (1994). Which whales are hunted? A molecular genetic approach to monitoring whaling. *Science* **265**, 1538–1539.

Baker, C.S., Lukoschek, V., Lavery, S., Dalebout, M.L., Yong-Un, M., Endo, T. & Funahashi, N. (2006). Incomplete reporting of whale, dolphin and porpoise 'bycatch'

revealed by molecular monitoring of Korean markets. *Anim. Conserv.* **9**, 474–482.

Barnes, R.F.W. (2002). The bushmeat boom and bust in West and Central Africa. *Oryx* **36**, 236–242.

Bensasson, D., Zhang, D.X., Hartl, D.L. & Hewitt, G.M. (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.* **16**, 314–321.

Bidlack, A.L., Reed, S.E., Palsbøll P, J. & Getz, W.M. (2007). Characterization of a western North American carnivore community using PCR-RFLP of cytochrome *b* obtained from fecal samples. *Conserv. Genet.* **8**, 1511–1513.

Bowkett, A.E., Plowman, A.B., Stevens, J.R., Davenport, T.R.B. & Jansen van Vuuren, B.J. (2008). Genetic testing of dung identification for antelope surveys in the Udzungwa Mountains, Tanzania. *Conserv. Genet.* **10**, 251–255.

Brashares, J.S., Arcese, P., Sam, M.K., Coppolillo, P.B., Sinclair, A.R.E. & Balmford, A. (2004). Bushmeat hunting, wildlife declines, and fish supply in West Africa. *Science* **306**, 1180–1183.

Chandiwana, S. & Ornbjerg, N. (2003). Review of north–south and south–south cooperation and conditions necessary to sustain research capability in developing countries. *J. Health Popul. Nutr.* **21**, 288–297.

Croes, B.M., Laurance, W.F., Lahm, S.A., Tchignoumba, L., Alonso, A., Lee, M.E., Campbell, P. & Buij, R. (2007). The influence of hunting on antipredator behavior in Central African monkeys and duikers. *Biotropica* **39**, 257–263.

Crooks, D.J., Ankudey, N. & Milner-Gulland, E.J. (2005). The value of a long-term bushmeat market dataset as an indicator of system dynamics. *Environ. Conserv.* **32**, 333–339.

DeYoung, R.W. & Honeycutt, R.L. (2005). The molecular toolbox: genetic techniques in wildlife ecology and management. *J. Wildl. Mgmt.* **69**, 1362–1384.

Dreher, B.P., Winterstein, S.R., Scribner, K.T., Lukacs, P.M., Etter, D.R., Rosa, G.J.M., Lopez, V.A., Libants, S. & Filcek, K.B. (2007). Noninvasive estimation of black bear abundance incorporating genotyping errors and harvested meat. *J. Wildl. Mgmt.* **71**, 2684–2693.

Dubost, G. (1980). Ecology and social-behavior of the blue duiker – small African forest ruminant. *Zeitschrift Fur Tierpsychologie – J. Comp. Ethol.* **54**, 205–266.

Dubost, G. (1984). Comparison of the diets of frugivorous forest ruminants of Gabon. *J. Mammal.* **65**, 298–316.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Ekrem, T., Willassen, E. & Stur, E. (2007). A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Mol. Phylogenet. Evol.* **43**, 530–542.

Ellis, A.M. & Bernard, R.T.F. (2005). Estimating the density of kudu (*Tragelaphus strepsiceros*) in subtropical thicket using line transect surveys of dung and distance software. *Afr. J. Ecol.* **43**, 362–368.

Fa, J.E., Juste, J., Burn, R.W. & Broad, G. (2002). Bushmeat consumption and preferences of two ethnic groups in Bioko Island, West Africa. *Human Ecol.* **30**, 397–416.

Fa, E.F., Seymour, S., Dupain, J., Amin, R., Albrechtsen, L. & Macdonald, D. (2006). Getting to grips with the magnitude of exploitation: bushmeat in the Cross-Sanaga rivers region, Nigeria and Cameroon. *Biol. Conserv.* **129**, 497–510.

Feer, F. (1989*a*). Comparative diet of *Cephalopus callipygus* and *C. dorsalis*, sympatric bovids of the African evergreen forest. *Mammalia* **53**, 563–604.

Feer, F. (1989*b*). The use of space by 2 sympatric duikers (*Cephalophus callipygus* and *Cephalophus dorsalis*) in an African rainforest: the role of activity rhythms. *Rev. D Ecol. – La Terre et La Vie* **44**, 225–248.

Focardi, S., Isotti, R. & Tinelli, A. (2002). Line transect estimates of artiodactyl populations in a Mediterranean forest. *J. Wildl. Mgmt.* **66**, 48–58.

Fong, J.J., Parham, J.F., Shi, H., Stuart, B.L. & Carter, R.L. (2007). A genetic survey of heavily exploited, endangered turtles: caveats on the conservation value of trade animals. *Anim. Conserv.* **10**, 452–460.

Food and Agricultural Organization. (1999). *Recognizing 'peste de petits ruminants'. A field manual.* Rome: United Nations.

Hammond, R.L., Macasero, W., Flores, B., Mohammed, O.B., Wacher, T. & Bruford, M.W. (2001). Phylogenetic reanalysis of the Saudi gazelle and its implications for conservation. *Conserv. Biol.* **15**, 1123–1133.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003). Biological identifications through DNA barcodes. *Proc. Roy. Soc. Lond. Ser. B* **270**, 313–321.

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S. & Francis, C.M. (2004). Identification of birds through DNA barcodes. *PLoS Biol.* **2**, 1657–1663.

Heydon, M.J. & Bulloh, P. (1997). Mouse deer densities in a tropical rainforest: the impact of selective logging. *J. Appl. Ecol.* **34**, 484–496.

Hickerson, M.J., Meyer, C.P. & Moritz, C. (2006). DNA barcoding will often fail to discover new animal species over broad parameter space. *Syst. Biol.* **55**, 729–739.

Hoelzel, A.R., Hancock, J.M. & Dover, G.A. (1991). Evolution of the cetacean mitochondrial D-loop region. *Mol. Biol. Evol.* **8**, 475–493.

Huelsenbeck, J.P. & Ronquist, F. (2001). MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

Jansen van Vuuren, B.J. & Robinson, T.J. (2001). Retrieval of four adaptive lineages in duiker antelope: evidence from mitochondrial DNA sequences and fluorescence in situ hybridization. *Mol. Phylogenet. Evol.* **20**, 409–425.

Karesh, W.B., Hart, J.A., Hart, T.B., House, C., Torres, A., Dierenfeld, E., Braselton, W.E., Pache, H. & Cook, R.A. (1995). Health evaluation of five sympatric duiker species (*Cephalophus* spp.). *J. Zoo Wildl. Med.* **26**, 485–502.

Kingdon, J. (1997). *The Kingdon field guide to African mammals*. London: Academic Press.

Kohn, M.H., York, E.C., Kamradt, D.A., Haught, G., Sauvajot, R.M. & Wayne, R.K. (1999). Estimating population size by genotyping faeces. *Proc. Roy. Soc. Lond. Ser. B* **266**, 657–663.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings Bioinform.* **5**, 150–163.

Lahm, S.A. (1993*a*) *Ecology and economics of human/wildlife interaction in northeastern Gabon*. PhD thesis, New York University.

Lahm, S.A. (1993*b*). Utilization of forest resources and local variations of wildlife populations in northeastern Gabon. In *Tropical forests, people and food: biocultural interactions and applications to development*: 213–227. Hladik, C.M., Hladik, A., Linares, O.F., Pagezy, H., Semple, A. & Hadley, M. (Eds). Paris: UNESCO.

Lahm, S.A., Kombila, M., Swanepoel, R. & Barnes, R.F.W. (2007). Morbidity and mortality of wild animals in relation to outbreaks of Ebola haemorrhagic fever in Gabon, 1994–2003. *Trans. Roy. Soc. Trop. Med. Hyg.* **101**, 64–78.

Lahm, S.A. & Tezi, J.-P. (2006). Assessment of the communities of medium-sized and large arboreal and terrestrial mammals in the Rabi/Toucan region of the Ngové-Ndogo hunting domain and southwestern Loango National Park. In *Gamba, Gabon: biodiversity of an equatorial African forest, bulletin of the biological society of Washington*: 383–417. Washington, DC: National Museum of Natural History.

Laurance, W.F., Croes, B.M., Tchignoumba, L., Lahm, S.A., Alonso, A., Lee, M., Campbell, P. & Ondzeano, C. (2006). Impacts of roads and hunting on Central African rainforest mammals. *Conserv. Biol.* **19**, 268–275.

Leaché, A.D. & Reeder, T.W. (2002). Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. *Syst. Biol.* **51**, 44–68.

Lorenz, J.G., Jackson, W.E., Beck, J.C. & Hanner, R. (2005). The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philos. Trans. Roy. Soc. Ser. B* **360**, 1869–1878.

Moodley, Y., Bruford, M.W., Bleidorn, C., Wronksi, T., Apio, A. & Plath, M. (2009). Analysis of mitochondrial DNA data reveals non-monophyly in the bushbuck (Tragelaphus scriptus) complex. *Mamm. Biol.* **74**, 418–422.

Moore, M.K., Bemiss, J.A., Rice, S.M., Quattro, J.M. & Woodley, C.M. (2003). Use of restriction fragment length polymorphisms to identify sea turtle eggs and cooked meats to species. *Conserv. Genet.* **4**, 95–103.

Moritz, C. & Cicero, C. (2004). DNA barcoding: promises and pitfalls. *PLoS Biol.* **2**, 1529–1531.

Newing, H. (2001). Bushmeat hunting and management: implications of duiker ecology and interspecific competition. *Biodivers. Conserv.* **10**, 99–118.

Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874–879.

Pääbo, S. & Wilson, A.C. (1988). Polymerase chain reaction reveals cloning artifacts. *Nature* **334**, 387–388.

Pilot, M., Gralak, B., Goszczynski, J. & Posluszny, M. (2007). A method of genetic identification of pine marten (*Martes martes*) and stone marten (*Martes foina*) and its application to faecal samples. *J. Zool. (Lond.)* **271**, 140–147.

Rambaut, A. & Drummond, A.J. (2007). *TRACER v1.4* Available at http://beast.bio.ed.ac.uk/Tracer.

Ross, H.A., Lento, G.M., Dalebout, M.L., Goode, M., Ewing, G., McLaren, P., Rodrigo, A.G., Lavery, S. and Baker, C.S. (2003). DNA surveillance: web based molecular identification of whales, dolphins and porpoises. *J. Hered.* **94**, 111–114.

Sambrook, J. & Russell, D.W. (2001). *Molecular cloning: a laboratory manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Shields, G.F. & Kocher, T.D. (1991). Phylogenetic relationships of North American Ursids based on analysis of mitochondrial DNA. *Evolution* **45**, 218–221.

Song, H., Buhay, J.E., Whiting, M.F. & Crandall, K.A. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc. Natl. Acad. Sci. USA* **105**, 13486–13491.

Swofford, D. (2000). *PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4*. Sunderland: Sinauer Associates.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L.P. & Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucl. Acids Res.* **24**, 3189–3194.

Taberlet, P. & Luikart, G. (1999). Non-invasive genetic sampling and individual identification. *Biol. J. Linn. Soc.* **68**, 41–55.

Thibault, M. & Blaney, S. (2003). The oil industry as an underlying factor in the bushmeat crisis in Central Africa. *Conserv. Biol.* **17**, 1807–1813.

Thompson, J.D., Gibson, T.J., Plewniak, G., Jeanmougin, F. & Higgins, D.G. (1997). The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* **25**, 4876–4882.

Travis, S.E. & Keim, P. (1995). Differentiating individuals and populations of mule deer using DNA. *J. Wildl. Mgmt.* **59**, 824–831.

Tutin, C.E.G., White, L.J.T. & Mackanga-Missandzou, A. (1997). The use by rain forest mammals of natural forest fragments in an equatorial African savanna. *Conserv. Biol.* **11**, 1190–1203.

van Vliet, N. (2008) *Variabilité spatiale et temporelle au sein système "chasseur – animal – territoire de chasse villageois" - pour une approche géographique de la durabilité de la chasse en Afrique Central*. PhD thesis, Universite Toulouse le Mirail, Toulouse, France.

van Vliet, N., Zundal, S., Miguel, C., Taberlet, P. & Nasi, R. (2008). Distinguishing dung from blue, red and yellow-backed duikers through noninvasive genetic techniques. *Afr. J. Ecol.* **46**, 411–417.

Vàzquez-Domínguez, E., Paetkau, C., Tucker, N., Hinten, G. & Moritz, C. (2001). Resolution of natural groups using iterative assignment tests: an example from two species of Australian native rats (*Rattus*). *Mol. Ecol.* **10**, 2069–2078.

Vrba, E.S. (1995). The fossil record of African antelopes (Mammalia, Bovidae) in relation to human evolution and paleoclimate. In *Palaeoclimate and evolution, with emphasis on human origins*: 385–424. Vrba, E.S., Denton, G.D., Partrige, T.C. & Burckle, L.H. (Eds). New Haven: Yale University Press.

Walsh, P.D. & White, L.J.T. (1999). What it will take to monitor forest elephant populations. *Conserv. Biol.* **13**, 1194–1202.

Waltert, M., Heber, S., Riedelbauch, S., Lien, J.L. & Muhlenberg, M. (2006). Estimates of blue duiker (*Cephalophus monticola*) densities from diurnal and nocturnal line transects in the Korup region, south-western Cameroon. *Afr. J. Ecol.* **44**, 290–292.

Wasser, S.K., Mailand, C., Booth, R., Mutayoba, B., Kisamo, E., Clark, B. & Stephens, M. (2006). Using DNA to track the largest ivory seizure since the 1989 trade ban. *Proc. Natl. Acad. Sci. USA* **104**, 4228–4323.

White, L. & Edwards, A. (2000). *Conservation en forêt pluviale africaine: Méthodes de recherche*. New York: Wildlife Conservation Society.

Wiens, J.J. & Penkroft, T.A. (2002). Delimiting species limits in spiny lizards (*Sceloporus*). *Syst. Biol.* **1**, 69–91.

Wilkie, D.S. & Carpenter, J.F. (1999). Bushmeat hunting in the Congo Basin: an assessment of impacts and options for mitigation. *Biodivers. Conserv.* **8**, 927–955.

Willows-Munro, S., Robinson, T.J. & Matthee, C.A. (2005). Utility of nuclear DNA intron markers at lower taxonomic levels: phylogenetic resolution among nine *Tragelaphus* spp. *Mol. Phylogenet. Evol.* **35**, 624–636.

Zapata, M.A., Cienfuegos, A.V., Ouirós, O.I., Ouiñones, M.L., Luckhart, S. & Correa, M.M. (2007). Discrimination of seven *Anopheles* species for San Pedro de Uraba, Antioquia, Columbia, by polymerase chain reaction-restriction fragment length polymorphism analysis of its sequences. *Am. J. Trop. Med. Hyg.* **77**, 67–72.

Zischler, H., Geisert, H., von Haeseler, A. & Paabo, S. (1995). A nuclear fossil of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**, 489–492.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Cytochrome *b* neighbour-joining bootstrap consensus phylogeny based on Kimura-2-parameter corrected distances and rooted with *H. aquaticus*. Sequences were aligned with CLUSTAL. Bootstrap values are indicated at the relevant node.

Figure S2. Cytochrome *b* maximum parsimony bootstrap consensus phylogeny rooted with *H. aquaticus*. Sequences were aligned with CLUSTAL. Bootstrap values are indicated at the appropriate node.

Figure S3. Control region phylogeny based on Bayesian analysis and rooted with *Tragelaphus* species. Sequences were aligned with CLUSTAL. Posterior probability values are indicated at the appropriate node.

Figure S4. Control region neighbour-joining bootstrap consensus phylogeny based on Kimura-2-parameter corrected distances and rooted with *N. moschatus*. Sequences were aligned with CLUSTAL. Bootstrap values are indicated at the relevant node.

Figure S5. Control region maximum parsimony bootstrap consensus phylogeny rooted with *N. moschatus*. Sequences were aligned with CLUSTAL. Bootstrap values are indicated at the appropriate node.

Figure S6. Flow chart illustrating the application of RFLP banding patterns to the diagnosis of species using cytochrome *b* gene sequence data.

Table S1. Details of the species, provider, geographic origin and GenBank accession numbers for samples used in the present study.

Table S2. Candidate cytochrome *b* gene RFLPs for diagnosing central African rainforest ungulate species and the size fragments (indicated in parentheses) generated by each enzyme.

Table S3. Candidate control region RFLPs for diagnosing central African rainforest ungulate species and the size fragments (indicated in parentheses) generated by each.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.