CHAPTER 2

MOLECULAR PHYLOGENY OF THE CEPHALOPHINI

OBJECTIVES

The aims of this aspect of the study were twofold. First, an attempt is made to determine evolutionary relationships between recognized duiker species using molecular sequence data from two mitochondrial DNA genes, cytochrome *b* and the small ribosomal subunit (12S rRNA). Secondly, by applying a molecular clock to date divergences, speciation within the tribe is correlated with biogeographical events which are known to have shaped forest biota.

INTRODUCTION

The assessment of evolutionary relationships have, traditionally, rested on morphological and phenotypic character states (Avise 1994). These characters are, however, frequently obscured by parallelisms and convergent evolution as are molecular data (Hedges & Maxson 1996) which may lead to spurious relationships. The tribe Cephalophini is a case in point. Several genera and species groups have been described using characters such as the length and texture of hair on the shoulders, the presence of inguinal glands, the length of the false hoof, and juvenile pelage patterns (Pocock 1910, St. Leger 1936, Heyden 1968, Groves & Grubb 1981; see Chapter 1). The inconsistency regarding the number of recognized species and genera within the tribe is therefore not surprising, and even today, relationships among species are largely unresolved.

The discovery of the polymerase chain reaction (PCR; Mullis & Faloona 1987) has had a profound effect on phylogenetic studies leading to the extensive use of DNA sequences to describe diversity and evolutionary relationships in all life forms. Arguably, the most frequently targeted sequences used to retrieve phylogenetic relationships are those derived from the mitochondrion. The extranuclear mitochondrial molecule is a closed, circular gene system that replicates and transcribes within the organelle itself (Avise & Lansman 1983). It has no intervening sequences, spacers, or repetitive DNA. Moreover, several unique

features make it especially useful as a tracer of macroevolutionary events. First, it is usually maternally inherited (Lansman *et al.* 1983, Gyllensten *et al.* 1985, Avise & Vrijenhoek 1987, but see Avise 1991, Gyllensten *et al.* 1991) which allows for an effective haploid transmission across generations. Secondly, the mitochondrial DNA molecule is free from recombination events (Harrison 1989). Thirdly, variation in the evolutionary rate of different parts of the molecule permits the selection of genes which are appropriate to address taxonomic questions that span a wide evolutionary time scale (Harrison 1991, Hillis & Dixon 1991).

The phylogenetic value of a gene, or region of DNA, rests on the premise that the rate of evolution matches the divergence times of the taxa under consideration (Hillis & Huelsenbeck 1992). The most informative studies compare sequences that vary in amount of divergence but are not so different that patterns are obscured by multiple events at the same nucleotide site (Moritz *et al.* 1987). In the case of the Bovidae, sequence data from the mitochondrial cytochrome *b* and 12S rRNA genes have been widely used to address evolutionary questions (Allard *et al.* 1992, Groves & Shields 1996, Robinson *et al.* 1996, Gatesy *et al.* 1997, Hassanin & Douzery 1999, Matthee & Robinson 1999a), and the same loci were selected in this study to investigate relationships between the cephalophine species. In so doing it was hoped that cytochrome *b* with its faster mutation rate would possibly resolve relationships near the terminal branches of the phylogenetic tree relative to the slower 12S rRNA characters which, based on other investigations (Moritz *et al.* 1987, Halanych & Robinson 1999), could be predicted to be distributed more basally.

MATERIAL AND METHODS

Sample collection and storage

Total genomic DNA was extracted from representatives of all 19 recognized duiker species (Grubb 1993; Table 2). In instances where the species concerned was characterized by an extensive range (i.e. *S. grimmia, C. monticola)* an attempt was made to include two or more specimens from widely separated localities. Since these specimens formed monophyletic clusters in preliminary investigations, only a single representative per species was used in the assessment of phylogenetic relationships. In these instances the retention of a particular specimen for subsequent analyses was based on its sharing the same geographic region as that of the holotype or, alternatively, that it was geographically the closest of the competing specimens to the holotype's locality.

Fresh material was the preferred source for DNA extractions (either soft tissue preserved in a 20% DMSO/saturated salt solution, or cells harvested from fibroblast cultures). Given the difficulties encountered in obtaining fresh samples, as well as the scarcity and endangered status of several of the taxa, teeth or dried skins were used as alternatives for DNA extraction *(C. ogilbyi, C. rubidus, C. zebra).* Table 2 lists the collection localities as well as type of material used for DNA extraction from all specimens included in the phylogenetic analysis.

Genomic DNA extraction and nucleotide sequencing

Total genomic DNA extraction from fresh material was done following standard protocols (Maniatis *et al.* 1982, Amos & Hoelzel 1991). These involved phenol/chloroform/iso-amyl alcohol steps with precipitation carried out overnight in ice-cold absolute ethanol. DNA pellets were re-suspended in ddH_2O to a final concentration of 1 μg/μl.

Hagelberg (1994) described an extraction protocol for DNA from ancient bones. This protocol, with minor modifications, was followed for DNA extractions from duiker teeth. A Rotary Power Tool (Sears Best/Craftsman) fitted with 0.05 mm drill-bits was used to drill into the pulp cavity of a molar and the resulting powder collected for DNA extractions. Incubation in a lysis buffer was overnight and the DNA was subsequently allowed to precipitate in ice-cold absolute ethanol for one week. Extractions using skin pieces obtained from museum specimens followed conventional protocols (as described for fresh material),

Table 2 Duiker species used in the present study. Vernacular names, collection localities, and tissues used fo DNA extractions are provided. Where multiple specimens were sequenced per species, those included in th phylogenetic analyses are indicated by an * (see text for details).

Species	Vernacular name	Locality	Source material	Material provided
*S. grimmia**	Gray duiker	Zimbabwe	Primary culture	V. Wilson[1]
S. grimmia	Gray duiker	CAR.	Soft tissue	E. Stockenstroom[2]
C. adersi	Ader's duiker	Zanzibar Island	Soft tissue	A. Williams[3]
*C. callipygus**	Peters' duiker	Gabon	Soft tissue	Aarn[4]
C. callipygus	Peters' duiker	Congo	Soft tissue	E. Stockenstroom[2]
C. dorsalis	Bay duiker	Ghana	Primary culture	V. Wilson[1]
C. harveyi	Harvey's red duiker	Tanzania	Soft tissue	L. Vinciguerra[5]
C. jentinki	Jentink's duiker	Brownsville Zoo.	Blood	Brownsville Zoo[6]
C. leucogaster	White-bellied duiker	Congo	Soft tissue	E. Stockenstroom[2]
*C. monticola**	Blue duiker	South Africa	Primary culture	V. Wilson[1]
C. monticola	Blue duiker	Cameroon	Soft tissue	P. Bishop[7]
C. monticola	Blue duiker	Gabon	Soft tissue	Aarn[4]
C. maxwellii	Maxwell's duiker	Ghana	Primary culture	V. Wilson[1]
C. natalensis	Natal red duiker	South Africa	Primary culture	V. Wilson[1]
C. niger	Black duiker	Ghana	Soft tissue	P. Arctander[8]
C. nigrifrons	Black-fronted duiker	Congo	Soft tissue	E. Stockenstroom[2]
C. ogilbyi	Ogilby's duiker	Nigeria	Dried skin	C. Powell[9]
C. silvicultor	Yellow-backed duiker	Philadelphia Zoo.	Primary culture	V. Wilson[1]
C. spadix	Abbot's duiker	Tanzania	Primary culture	L. Vinciguerra[5]
C. rubidus	Ruwenzori red duiker	Ruwenzori Mt.	Tooth	S. Strand[10]
C. rufilatus	Red-flanked duiker	CAR.	Soft tissue	E. Stockenstroom[2]
C. weynsi	Weyne's duiker	Rwanda	Soft tissue	C. Claude[11]
C. zebra	Zebra duiker	Liberia	Tooth	C. Claude[11]

CAR=Central African Republic; 1=Chipangali Wildlife Orphanage, Zimbabwe; 2=Congo Safaris, Sou Africa; 3=Zanzibar Protected Areas Project, Commission for Natural Resources, Tanzania; 4=Aarn, Australi 5=Safariroyal, Tanzania; 6=Brownsville Zoological Garden, USA; 7=University of the Witwatersrand, Sou Africa; 8=University of Copenhagen, Denmark; 9=University of Port Harcourt, Nigeria; 10=Swedish Museu of Natural History, Sweden; 11=Zoologisches Museum, Universität Zürich, Switzerland.

however, the incubation and precipitation steps were prolonged and were similar to tooth extractions. In view of the possibility of contamination when working with degraded DNA typically retrieved from museum specimens, procedures routinely used in the laboratory were followed (Robinson *et al.* 1996, Matthee & Robinson 1999b, Robinson & Matthee 1999). Among others, these included the DNA extractions being performed in a separate DNA-free laboratory, multiple extractions, the amplification and sequencing of the museum specimens being separated temporally, and that the sequence data made phylogenetic sense (Austin *et al.* 1997).

Two mitochondrial DNA genes were targeted for phylogenetic inference. Cytochrome *b* (full gene) and 12S rRNA (approximately 800 bp on the 5' side of the subunit) were amplified by PCR using universal primers (cytochrome *b*: Pääbo & Wilson 1988, Kocher *et al.* 1989, Irwin *et al.* 1991, Matthee & Robinson 1999a; 12S rRNA: Allard & Honeycutt 1992). To allow for single stranded sequencing, end primers were biotin coated. PCR reactions were set up in a laminar flow hood to minimize the risk of contamination and negative controls, comprising all components except DNA, were always included. Given the degraded nature of DNA extracted from museum material, amplification was not always successful when using universal primers that spanned sections larger than 350 bp. Customized internal primers were therefore designed for both amplification and sequencing. Table 3 provides a summary of PCR and sequencing primers used in this study.

A 35 cycle PCR program was routinely followed (initial denaturation for 3 min at 94°C, denaturation for 45 sec at 94°C, annealing for 45 sec at 50°C, and extension for 1 min at 72°C with a final extension cycle of 5 min at 72°C). PCR products were separated in 1% agarose gels (Techcomp LTD), and fragments extracted using Nucleotrap Extraction Kit for Nucleic Acids (Macherey-Nagel).

Sequencing followed Sanger *et al.* (1977) using the Sequenase Kit v.2.0 (United States Biochemical Corporation) with [α-^{32}P] dATP as the labeling agent. Single stranded template was obtained by treating the purified product with dynabeads-streptavidin (Dynal A.S.), a magnetic particle with an affection for biotin, followed by a denaturation step involving 0.1 M NaOH. Overlap in nucleotide sequences obtained from L and H sequencing primers provided verification of sequences.

Table 3 Primers used and their sequences for both amplification and sequencing of duiker DNA. Biotin coated end primers are indicated by *. Customized primer numbers correspond to *Bos taurus* (Anderson *et al.* 1982) and the sequences are reported 5' to 3'.

Cytochrome b	Reference
*L14724	Pääbo & Wilson 1988
L14841	Kocher *et al.* 1989
L15162	Pääbo & Wilson 1988
L15408	Irwin *et al.* 1991
BL	Matthee & Robinson 1999a
AH	Matthee & Robinson 1999a
H15149	Kocher *et al.* 1989
*H15494	Irwin *et al.* 1991
*H15915	Pääbo & Wilson 1988

12S rRNA	Reference
*L82 (A)	Allard & Honeycutt 1992
L509 (C)	Allard & Honeycutt 1992
L673	AGCCACCGCGGTCATACG
L941	CTCAAAGGACTTGGCGGTG
H708	CACGTTTTACGCCGTATTCC
H618 (B)	Allard & Honeycutt 1992
H975	TTATAGAACAGGCTCCTC
*H900 (D)	Allard & Honeycutt 1992

Sequence analysis

The inclusion of more than one outgroup taxon for phylogenetic analyses may limit the introduction of errors in the polarization of characters and, by implication, the retrieval of erroneous evolutionary relationships (Milinkovitch *et al.* 1996, Milinkovitch & Lyons-Weiler 1998). To test the robustness of ingroup topologies, and also to avoid possible biases introduced by using a single outgroup, three outgroup taxa of successive relatedness to the ingroup were selected. These comprised the klipspringer, *O. oreotragus* (tribe Neotragini: cytochrome *b;* Matthee & Robinson 1999a; 12S rRNA; Hassanin & Douzery 1999), cattle, *B. taurus* (tribe Bovini: Anderson *et al.* 1982) and the more distantly related pronghorn antelope, *Antilocapra americana* (Family Antilocapridae: cytochrome *b;* Irwin *et al.* 1991; 12S rRNA; Kraus & Miyamoto 1991). For illustrative purposes trees incorporating all three outgroups simultaneously are presented. It is noteworthy, however, that although some minor branch swapping occurred when the outgroups were used singly, the main phylogenetic conclusions remained unchanged.

Both cytochrome *b* and 12S rRNA gene sequences were aligned by eye. The cytochrome *b* sequences were translated into protein codons. Both nucleotide sequences and their corresponding amino acids were examined for irregularities (e.g. stopcodons in the reading frame and that the nucleotide substitution pattern followed that of conventional protein coding genes) which could indicate nonfunctional nuclear homologs (see Arctander 1995, Zhang & Hewitt 1996 and references therein). All nucleotide sequences resulting in a unique amino acid change were verified. The 12S rRNA segment was assembled against the secondary structure model for *B. taurus* (Springer *et al.* 1995). When aligning the 12S rRNA sequences for the ingroup taxa, a single 3 bp insertion was found in the variable loop region between stem positions 35 and 36 (see Springer *et al.* 1995) which was common to both *C. monticola* and *C. maxwellii*. Sequence alignment to the outgroup taxa revealed several other insertions (between 1 and 3 bp in length) but these were unique to the outgroup sequences (i.e. were not present in the duiker antelope). Since the aim was to determine ingroup relationships, these indels were deleted from the outgroup species in all phylogenetic analyses. Sequences generated in the present study were deposited in Genbank (accession numbers AF153883-AF153905, AF154247-AF154270), and the alignments will be available in EMBL Treebase (http://phylogeny.harvard.edu/treebase).

Saturation of nucleotide changes within the Cephalophini was determined by plotting the actual number of transitions and transversions against time, expressed as sequence

divergence. Since corrected divergences might be more linear with respect to time than uncorrected values (Hackett 1996), only uncorrected p-distances are shown. Moreover, the use of uncorrected p-distances further allowed for comparisons of saturation with other bovid tribes. Given that different portions of the genes evolve at different rates, and would thus exhibit different saturation curves, the three codon positions were plotted separately for cytochrome *b* as well as stem and loop changes in the 12S rRNA gene.

The effects of deviations from a molecular clock on phylogeny reconstruction is well documented (DeBry 1992, Takezaki *et al.* 1995, Robinson *et al.* 1998). The relative rates of evolutionary change for individual lineages within the tribe Cephalophini were determined by RRTree (Robinson *et al.* 1998). Since the amount of noise being introduced on the branch leading to the reference taxon could influence the results, the closest relative outgroup *(O. oreotragus)* was used as reference taxon. This test was chosen because it takes the phylogenetic branching pattern into account and corrects for unbalanced sampling by assigning weights proportionally to branches. Given the differences in evolutionary rates between synonymous and non-synonymous substitutions, these two classes were separated when calculating evolutionary rates.

The data were analyzed by distance (neighbor joining algorithm; Saitou & Nei 1987), parsimony, and maximum likelihood (Felsenstein 1981) using Paup 4.0b2a (Swofford 1999). The HKY correction model (Hasegawa *et al.* 1985) was applied to the distance matrix for neighbor joining analyses. This correction takes the unequal base frequencies observed for most mitochondrial DNA genes into account and adjusts for different rates of transition and transversion accumulation. To accommodate among site rate variation, gamma shape parameters were calculated *via* maximum likelihood in Paup 4.0b2a using the optimal parsimony trees as the reference topology (shape parameter for cytochrome *b*=0.168; shape parameter for 12S rRNA=0.164). Distance and maximum likelihood trees constructed with an adjustment for among site rate variation were compared to those constructed under the assumption of equal rates.

Parsimony trees were generated using the heuristic search option in Paup 4.0b2a with 100 random replicates. To account for the transitional bias, all analyses were carried out using unweighted characters, transversion (tv) weighted at 2:1 over transitions (ti), and the empirical ti/tv ratio for the ingroup (13:1 for cytochrome *b*; 7:1 for 12S rRNA). Empirical values were calculated in MacClade (Maddison & Maddison 1992) by counting the average number of changes over 1 000 random trees (Halanych 1996). In the case of maximum

likelihood analyses, the empirical ti/tv ratios were specified for the genes and base frequencies were calculated from the data.

The skewness of the tree-length distribution (g1; Hillis & Huelsenbeck 1992), an indication of the phylogenetic signal present in the data, was calculated for 10 000 random trees (Paup 4.0b2a). Since closely related taxa can contribute to the amount of overall signal, g1 values were estimated for subsets of the data excluding closely related taxa.

Without an appraisal of the support for a particular clade, a phylogeny is of limited value (Sanderson 1995). Although not without its detractors (e.g. Faith & Cranston 1991, Kluge & Wolf 1993, Sitnikova *et al.* 1995), the nonparametric bootstrap technique (Efron 1979, Felsenstein 1985) is the most frequently used statistical method to assess the reliability of specific groupings of taxa. Bootstrapping entails random resampling of characters with replacement and its usefulness stems, in part, from the fact that it can be used in conjunction with many different tree-building algorithms and types of data. To assess the robustness of topologies retrieved by this study one thousand bootstrap replicates were performed for distance and parsimony analyses; due to computational difficulties, however, only 100 replicates were done for maximum likelihood.

In contrast to conventional (nonparametric) bootstrapping where biases present in the original data matrix are retained in the pseudoreplicates, the parametric bootstrap creates independent replicates of a data matrix based on numerical simulations using parameters estimated from the original data (Efron 1985, Felsenstein 1988, Huelsenbeck *et al.* 1996). In addition to this attribute, parametric bootstrapping is also useful for examining possible branch-length problems that may result in artificial relationships (Huelsenbeck *et al.* 1996). Simulated data sets were created with the Siminator computer program (Huelsenbeck *et al.* 1996) using the cytochrome *b* parsimony topology as the model tree. One thousand simulated data sets were generated under an HKY model of sequence evolution, with gamma shape parameter (0.168), ti/tv ratio (kappa=27.167) and nucleotide frequencies estimated by maximum likelihood (Paup 4.0b2a). Tree reconstructions were performed in Paup 4.0b2a using parameters identical to those adopted in the cytochrome *b* parsimony analyses with the exception of an unordered weighting scheme and the "as-is" sequence addition as opposed to 100 random additions. Since the specification of a model tree is a prerequisite for simulating data matrices, the largely unresolved topologies obtained from the small ribosomal sequence data were considered inadequate as input models for parametric bootstrap analysis of this gene.

All cytochrome *b* analyses were carried out using the complete sequence for the gene. For the small ribosomal subunit, compensatory changes in stem regions have led various authors (Dixon & Hillis 1993, Springer *et al.* 1995) to down-weight paired characters. In the parsimony analyses of this study, stem characters were assigned a base weight of 0.6 relative to loop characters. The 3 bp indel found in the conservative dwarf lineage was treated as an additional single character state in parsimony analyses (characters 701 and 702 were excluded from the data matrix and only character 703 was included in analyses); when applying different weighting schemes for transitions and transversions, the indel was assigned the same weight as a transition (Jansen van Vuuren & Robinson 1997).

Several contrasting ideas have been proposed on how best to analyze multiple data sets (Bull *et al.* 1993, Huelsenbeck *et al.* 1994, De Queiroz *et al.* 1995). In the present investigation, the cytochrome *b* and 12S rRNA data sets were analyzed separately and, since the partition homogeneity test (Farris *et al.* 1995 in Paup 4.0b2a) resulted in no significant heterogeneity (P=0.226) between them, they were subsequently combined (Hillis 1987, Swofford 1991). For the combined parsimony analysis, the empirical ti/tv ratios for the two genes were retained, stem characters were down-weighted by 0.6, and the indel was treated as a fifth character in 12S rRNA. For maximum likelihood analyses, the combined empirical ti/tv ratio (11:1) was specified.

RESULTS AND DISCUSSION

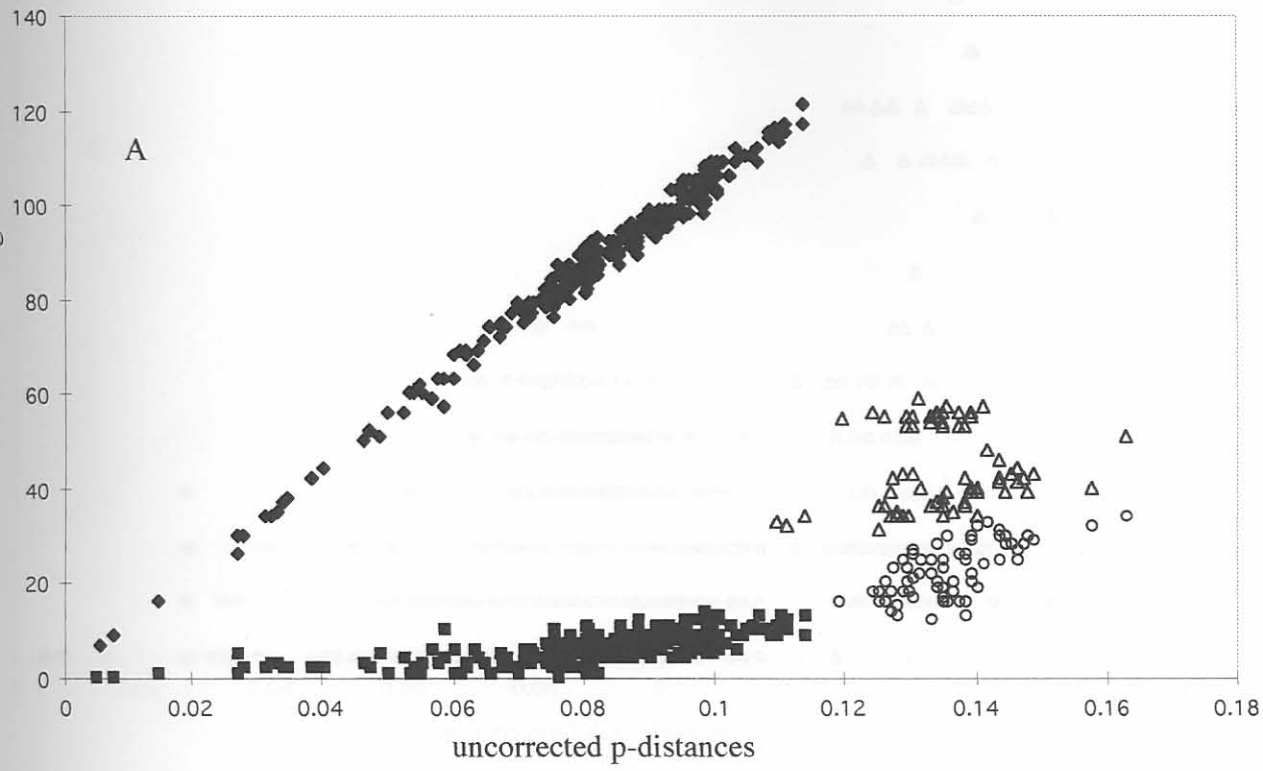
Saturation analysis

No evidence of saturation was found for cytochrome *b* when the number of transitions and transversions were plotted against time (expressed as uncorrected sequence divergences; Fig. 1A). This is an interesting finding since there are reports, encompassing several mammalian families, of saturation occurring as sequence divergences approach ≥ 10%; this includes the Bovidae (Irwin *et al.* 1991, Griffiths 1997, Matthee & Robinson 1999a). In contrast, although several of the duiker species are separated by sequence divergences similar to this (the highest uncorrected p-distance between duiker species being 11.4% between *C. monticola* and *C. harveyi;* see Table 4) there was no evidence of saturation in the Cephalophini data set. The linear increase evidenced by the plots may reflect the low accumulation of transversions in the ingroup (see Fig. 1). When different codon positions are considered separately, no saturation was observed for either first (Fig. 1B) or second codon changes (Fig. 1C). Third codon transitions showed some leveling off for certain ingroup comparisons (Fig. 1D), mostly involving either *C. monticola* or *C. maxwellii*. However, the majority of changes are linear with time, thus indicating the general absence of saturation at the third codon position. The lack of any distortion introduced by third codon position substitutions is also evident when CI values are determined for all ingroup changes (CI=0.431) versus only third codon changes (CI=0.421).

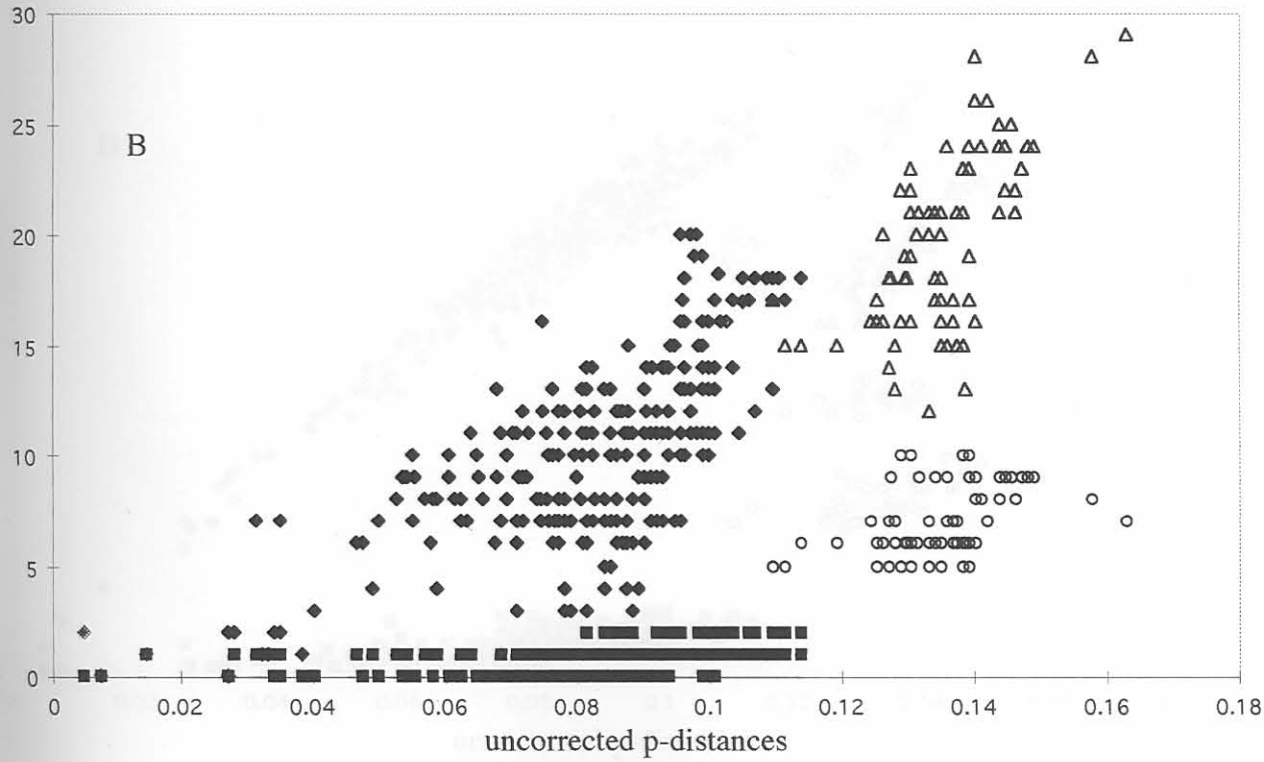
In the case of the 12S rRNA data, the highest uncorrected sequence divergence was 5.7% (between *C. maxwellii* and *C. nigrifrons;* see Table 5); not unexpectedly therefore, pairwise comparisons for this gene showed a monotonic linear increase over time (Fig. 1E). Considering paired and unpaired regions separately, loop characters may be expected to show some saturation given the weak evolutionary constraints acting on these domains (Springer *et al.* 1995). However, no evidence of this was found in the loop regions, either for transitions or transversions (Fig. 1F). Additionally, the CI values computed from the complete gene sequences obtained from the ingroup taxa versus those from the loop domain alone were similar (CI for complete data set=0.597; CI for loop domain=0.597). Given the conservative nature of the stems, the CI value for this region was, as expected, considerably higher (CI for stem domain=0.789).

Figure 1 Saturation plots for all duiker taxa included in the present study. The actual number of transitions and transversions are plotted against uncorrected p-distances. Closed symbols indicate ingroup comparisons while open symbols represent ingroup to outgroup comparisons. (A) Complete cytochrome *b* gene (B) Cytochrome *b* first codon position (C) Cytochrome *b* second codon position (D) Cytochrome *b* third codon position (E) Complete 12S rRNA gene (F) 12S rRNA loop domain.

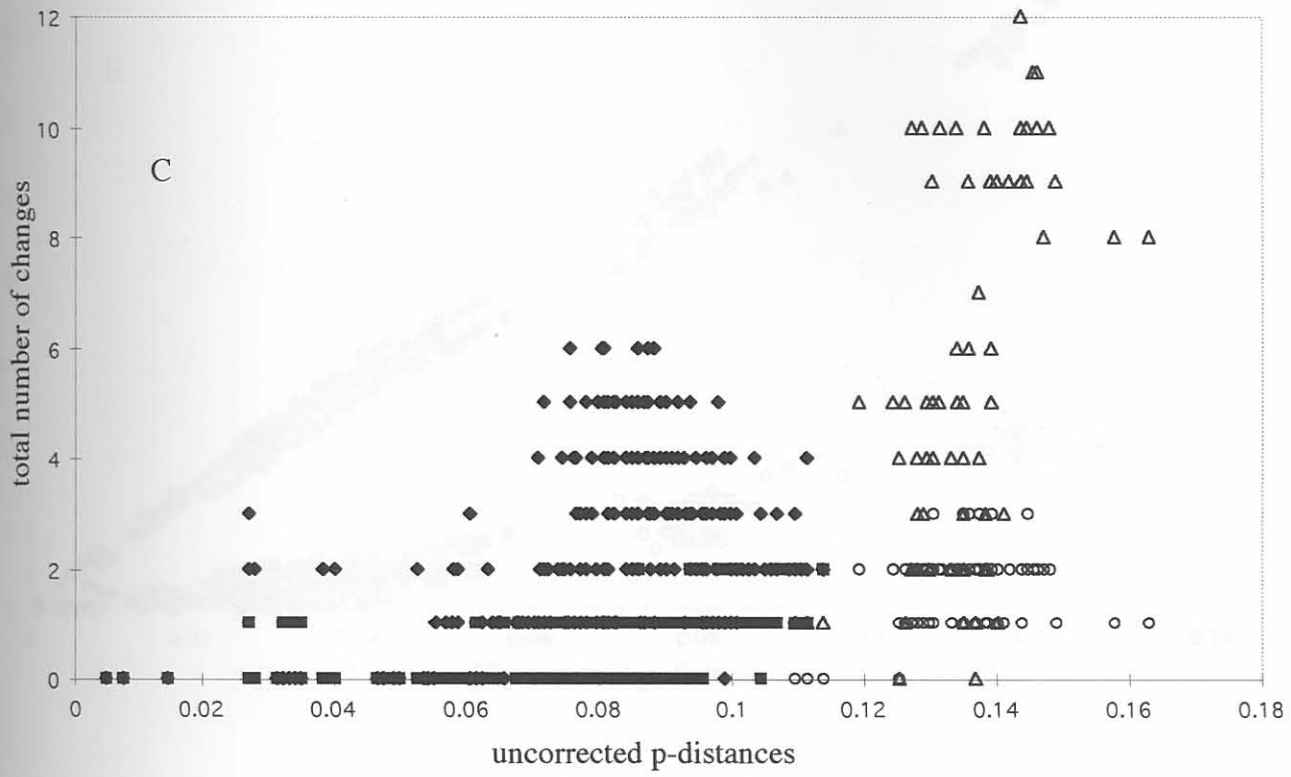
Complete cytochrome *b* gene



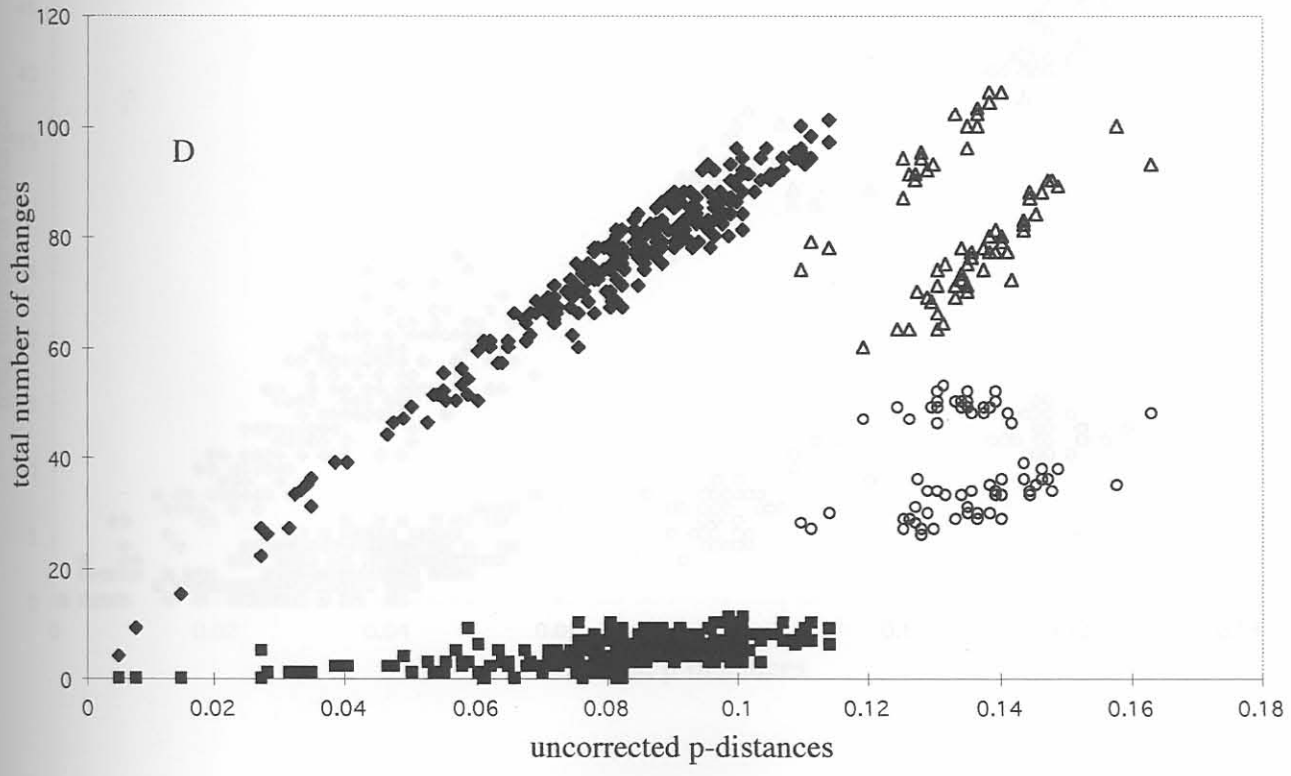
1st codon position

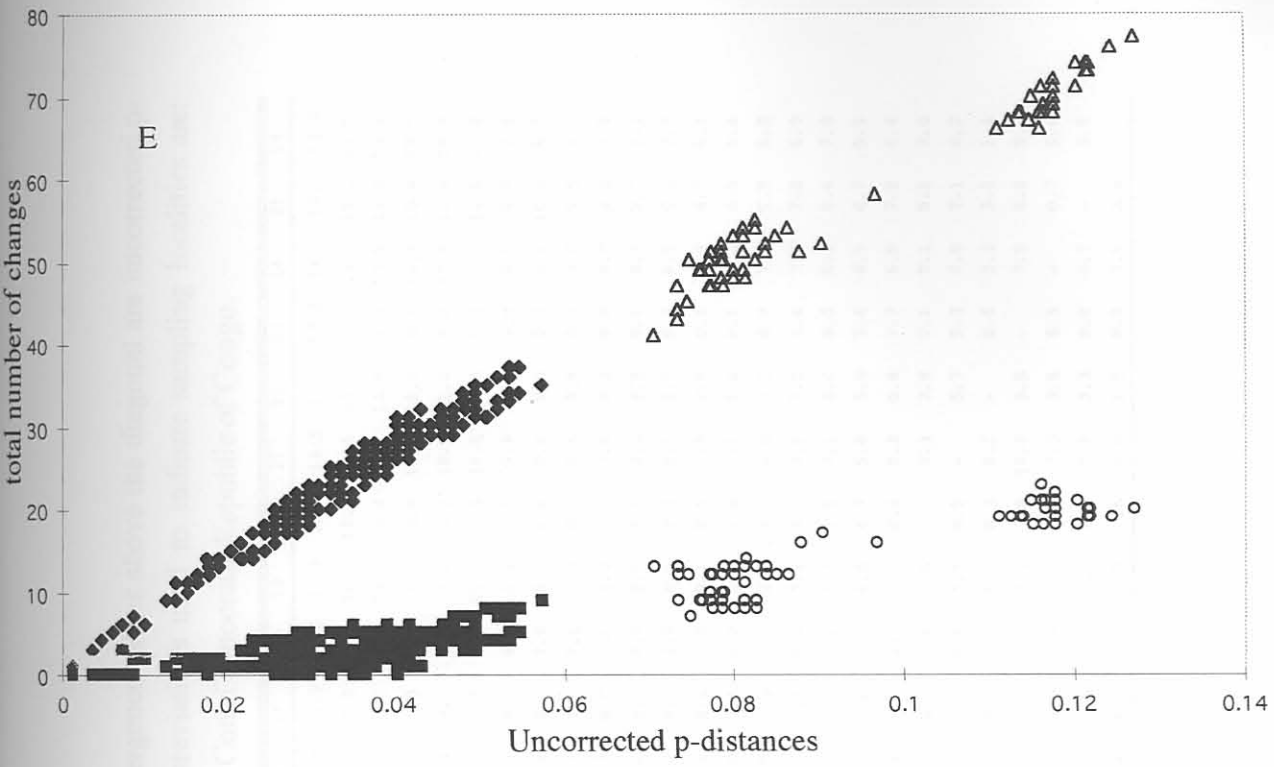


2nd codon position



3rd codon position





12S rRNA loop characters

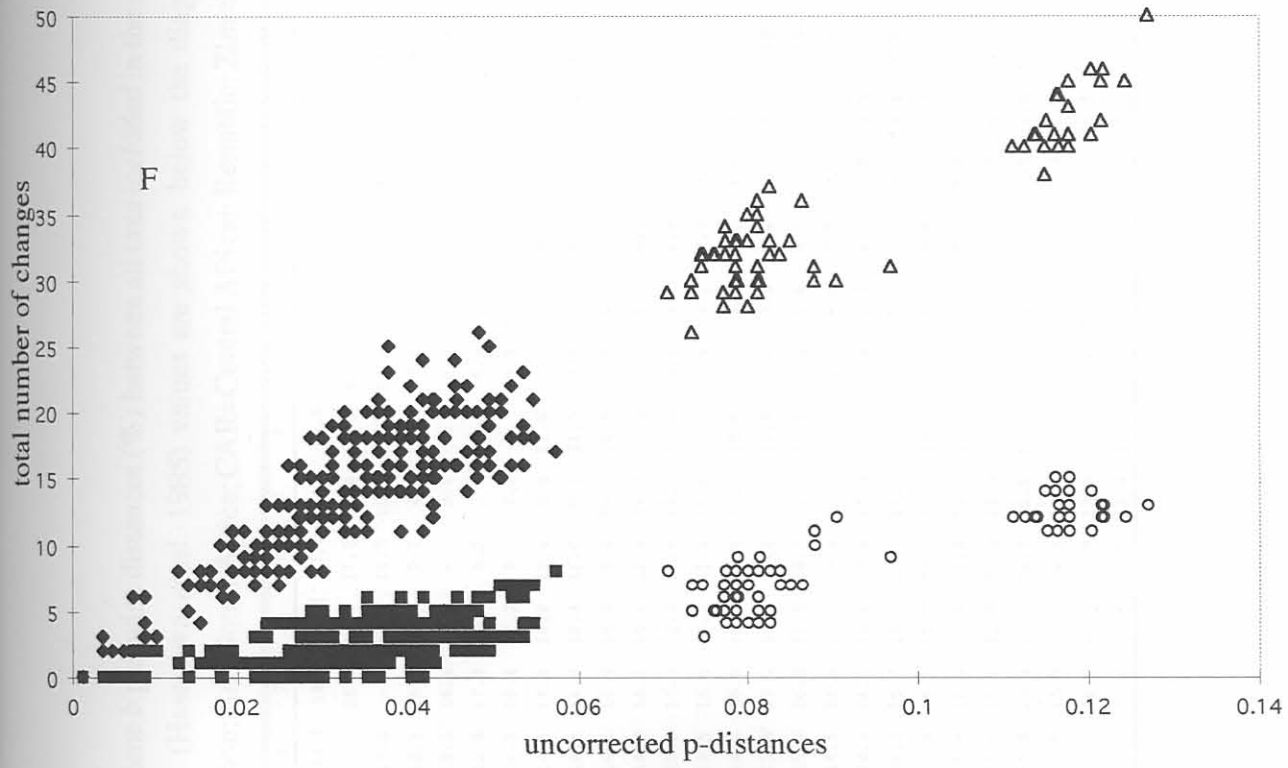


Table 4 Observed cytochrome *b* pairwise distances (%) between all taxa included in the study. Divergence values above the diagonal are uncorrected p-distances, while the HKY (Hasegawa *et al.* 1985) values are shown below the diagonal. Abbreviations used to indicate sampling localities are: Cam=Cameroon; Gab=Gabon; SA=South Africa; CAR=Central African Republic; Zim=Zimbabwe; Con=Democratic Republic of Congo.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Antilocapra	-	14.1	14.2	13.3	13.0	13.0	13.8	13.7	13.5	13.1	13.5	13.4	13.9	12.9	13.9	13.4	11.9	12.6	12.4	12.9	13.5	13.3	13.0	13.5	13.7	13.0
Oreotragus	15.9	-	16.3	10.9	11.1	11.4	12.5	12.8	12.8	12.5	12.7	13.5	12.8	12.6	13.8	13.5	12.8	13.0	12.7	13.3	13.6	13.6	13.8	13.6	14.0	13.5
Bos	16.1	19.0	-	14.0	14.3	14.9	15.7	12.8	13.0	14.3	14.6	14.5	14.6	14.0	14.8	14.4	12.7	13.1	13.8	13.4	14.7	13.9	14.3	13.9	14.4	13.5
monticola(Cam)	14.9	12.1	15.9	-	2.7	4.9	6.7	9.2	9.3	9.1	9.2	10.3	9.2	9.6	9.9	9.9	9.0	9.5	9.5	9.6	10.0	10.6	9.1	10.7	10.8	10.5
monticola(Gab)	14.6	12.3	16.4	2.7	-	5.8	7.7	9.3	9.9	9.8	10.0	11.1	10.0	9.5	9.7	9.3	9.0	9.4	9.2	9.2	10.0	11.1	8.8	10.8	11.4	10.9
monticola(SA)	14.5	12.6	17.1	5.1	6.2	-	7.1	9.3	8.5	8.8	8.8	10.0	9.6	10.0	10.7	10.7	9.3	9.8	10.0	9.9	10.9	10.5	9.9	11.0	11.4	10.2
maxwelli	15.6	14.1	18.4	7.2	8.3	7.6	-	10.3	10.0	9.7	10.0	10.9	9.9	9.7	10.1	10.0	9.1	9.7	9.4	9.5	9.9	9.9	9.0	9.5	9.8	9.6
grimmia(CAR)	15.5	14.4	14.4	10.2	10.4	10.4	11.6	-	2.7	8.5	9.2	8.2	8.6	8.2	9.0	9.0	8.0	7.6	8.1	8.4	9.4	9.0	9.0	9.6	10.0	8.7
grimmia(Zim)	15.2	14.4	14.6	10.4	11.0	09.4	11.3	2.8	-	8.2	8.4	7.4	8.8	7.9	8.7	8.7	7.7	7.6	7.8	8.5	9.5	8.9	8.7	9.0	9.5	8.5
silvicultor	14.6	14.1	16.4	10.0	10.9	09.7	10.8	9.3	9.0	-	2.8	6.0	7.1	8.5	8.5	8.7	8.0	7.5	8.0	8.0	7.6	8.2	8.2	9.0	9.2	7.8
spadix	15.1	14.3	16.7	10.2	11.1	09.7	11.3	10.2	9.2	2.8	-	5.7	7.7	9.0	9.1	9.2	8.7	8.0	8.5	8.5	9.0	8.5	8.1	9.3	9.7	8.1
dorsalis	15.0	15.3	16.7	11.6	12.5	11.2	12.3	9.0	8.0	6.4	6.0	-	7.1	8.5	8.9	8.7	8.0	8.0	8.5	8.4	8.5	8.1	8.7	9.1	9.2	7.8
jentinki	15.7	14.5	16.8	10.2	11.1	10.6	11.0	9.5	9.7	7.7	8.4	7.6	-	8.8	8.1	7.8	8.2	7.5	8.1	8.5	7.8	8.8	0.4	8.9	8.7	8.4
leucogaster	14.4	14.2	16.0	10.7	10.6	11.2	10.8	9.0	8.7	9.3	9.9	9.3	9.7	-	8.2	8.0	7.4	6.7	7.5	7.6	7.1	5.4	8.1	6.1	6.5	5.5
callipygus(Gab)	15.7	15.8	17.0	11.0	10.8	12.0	11.4	10.0	9.6	9.4	10.0	9.8	8.9	9.0	-	0.5	7.5	5.3	3.5	6.0	4.0	7.1	8.7	7.7	7.9	6.9
callipygus(Con)	15.0	15.3	16.5	11.0	10.3	12.0	11.1	10.0	9.6	9.6	10.2	9.6	8.4	8.8	0.5	-	7.3	5.0	3.1	5.5	3.8	7.1	8.4	7.5	7.8	6.9
zebra	13.1	14.5	14.2	10.0	10.0	10.3	10.1	8.8	8.4	8.8	9.5	8.7	9.0	8.1	8.2	8.0	-	7.2	6.3	7.8	7.1	8.0	8.0	8.2	8.6	7.8
rubidus	14.0	14.7	14.8	10.6	10.4	10.9	10.8	8.3	8.3	8.1	8.8	8.8	8.2	7.3	5.7	5.3	7.8	-	4.6	4.7	5.8	5.8	7.4	6.5	6.7	5.5
ogilbyi	13.8	14.3	15.7	10.5	10.1	11.2	10.5	8.9	8.6	8.7	9.4	9.4	8.8	8.2	3.6	3.2	6.7	4.9	-	6.2	5.2	6.4	7.7	6.8	7.2	6.4
niger	14.4	15.1	15.1	10.7	10.1	11.0	10.6	9.2	9.3	8.8	9.4	9.2	9.4	8.3	6.4	5.8	8.5	5.0	6.6	-	7.4	7.6	7.1	8.1	8.2	7.0
weynsi	15.1	15.6	16.9	11.1	11.2	12.3	11.0	10.5	10.6	8.2	9.9	9.4	8.4	7.6	4.2	4.0	7.7	6.1	5.5	8.1	-	5.7	9.3	6.8	7.1	6.2
rufilatus	14.9	15.6	15.8	11.9	12.6	11.8	11.0	10.0	9.8	9.0	9.4	8.9	9.7	5.7	7.7	7.7	8.8	6.2	6.9	8.3	6.1	-	8.6	3.3	3.2	1.4
adersi	14.6	15.8	16.4	10.0	09.7	11.0	09.9	10.0	9.6	9.0	8.9	9.6	11.7	8.9	9.6	9.2	8.8	8.1	8.4	7.8	10.4	9.5	-	7.8	8.2	8.0
natalensis	15.1	15.6	15.8	12.0	12.2	12.4	10.6	10.7	10.0	9.9	10.3	10.0	9.8	6.5	8.4	8.2	9.0	6.9	7.3	8.9	7.3	3.4	8.5	-	0.7	3.4
harveyi	15.4	16.0	16.5	12.2	12.9	12.9	10.9	11.2	10.6	10.1	10.8	10.3	9.6	7.0	8.7	8.5	9.5	7.3	7.8	9.0	7.6	3.3	9.0	0.7	-	3.5
nigrifrons	14.5	15.3	15.4	11.8	12.4	11.4	10.7	9.6	9.3	8.4	8.9	8.6	9.2	5.8	7.4	7.4	8.5	5.8	6.8	7.6	6.6	1.5	8.8	3.5	3.6	-

Table 5 Pairwise distances (%) calculated for the small ribosomal subunit. Divergence values above the diagonal line are uncorrected p-distances while HKY (Hasegawa *et al.* 1985) values are shown below the diagonal line. Abbreviations used are: Cam=Cameroon; Gab=Gabon; SA=South Africa; Zim=Zimbabwe; Con=Democratic Republic of Congo

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Antilocapra	-	12.9	13.6	12.2	12.1	12.1	12.5	13.0	13.0	12.9	13.7	13.3	12.7	12.7	12.3	12.1	12.8	12.5	12.5	12.4	12.5	12.8	12.5	12.9	13.2	12.5
Oreotragus	14.4	-	13.0	11.6	11.5	11.3	11.1	10.4	10.2	11.2	11.0	11.0	10.4	11.1	11.1	11.0	11.2	11.2	10.2	11.2	10.7	11.9	11.2	11.9	11.9	11.9
Bos	15.5	14.6	-	7.8	7.9	7.9	7.7	7.7	7.6	7.6	8.2	8.3	8.3	7.8	8.1	8.2	7.9	8.0	7.9	8.2	7.8	7.9	7.3	7.8	7.9	7.7
monticola(Cam)	13.8	12.8	8.4	-	0.3	1.4	2.3	5.0	5.0	5.2	5.3	5.0	4.8	4.4	4.3	4.4	4.4	4.3	4.4	4.8	3.7	4.9	4.1	5.1	5.4	5.2
monticola(Gab)	13.6	12.7	8.6	0.3	-	1.8	2.7	4.7	4.7	5.0	5.2	4.9	4.7	4.3	3.9	4.0	4.3	4.1	4.3	4.7	3.6	4.8	4.3	4.9	5.3	5.1
monticola(SA)	13.6	12.5	8.6	1.4	1.8	-	1.9	4.7	4.9	5.1	5.4	4.7	4.7	4.0	4.1	4.3	4.0	4.1	4.3	4.9	3.6	4.8	4.0	4.9	5.3	5.1
maxwellii	14.1	12.2	8.3	2.4	2.8	1.9	-	4.8	4.5	5.4	5.6	5.1	4.8	4.4	4.5	4.7	4.7	4.8	4.7	5.1	4.0	5.4	4.4	5.6	5.4	5.7
grimmia(Cam)	14.8	11.4	8.3	5.3	4.9	4.9	5.0	-	1.3	4.1	3.7	4.0	4.1	4.0	3.5	3.6	4.5	3.5	3.6	3.7	3.2	4.4	3.7	4.3	4.1	4.4
grimmia(Zim)	14.8	11.0	8.2	5.3	4.9	5.2	4.7	1.3	-	3.9	3.5	3.5	3.6	4.0	3.2	3.4	4.3	3.7	3.1	3.7	3.0	4.7	3.7	4.5	4.4	4.7
silvicultor	14.7	12.3	8.2	5.5	5.3	5.3	5.8	4.3	4.0	-	1.9	2.7	2.6	3.0	3.0	3.1	3.5	3.0	2.8	2.7	2.7	3.9	3.7	3.2	3.6	3.9
spadix	15.7	12.0	8.9	5.6	5.5	5.8	5.9	3.9	3.6	2.0	-	2.6	2.7	3.9	3.1	3.2	3.6	3.4	2.7	3.1	3.1	4.3	4.1	3.9	4.0	4.3
dorsalis	15.2	12.0	9.1	5.3	5.2	4.9	5.3	4.2	3.6	2.8	2.6	-	2.7	3.4	2.8	3.0	3.4	3.1	2.7	3.4	2.6	3.7	3.6	3.6	3.7	3.7
jentinki	14.3	11.4	9.1	5.0	4.9	4.9	5.0	4.3	3.8	2.6	2.8	2.8	-	3.2	2.4	2.3	3.0	2.7	1.8	2.4	2.2	4.1	3.7	3.5	3.4	4.1
leucogaster	14.3	12.2	8.5	4.6	4.4	4.2	4.6	4.2	4.2	3.1	4.0	3.5	3.3	-	2.3	2.4	3.0	1.8	2.4	2.8	2.0	2.6	3.1	2.3	2.7	2.8
callipygus(Gab)	13.8	12.2	8.8	4.5	4.0	4.3	4.7	3.6	3.3	3.1	3.2	2.9	2.5	2.4	-	0.1	2.3	1.7	0.9	1.8	0.5	3.0	2.3	2.6	2.7	3.2
callipygus(Con)	13.6	12.0	8.9	4.6	4.2	4.5	4.9	3.8	3.5	3.2	3.3	3.1	2.4	2.5	0.1	-	2.2	1.8	0.7	1.9	0.6	3.1	2.4	2.7	2.8	3.4
zebra	14.5	12.3	8.6	4.6	4.5	4.2	4.9	4.8	4.5	3.6	3.8	3.5	3.1	3.1	2.4	2.2	-	3.2	1.7	3.6	2.3	3.5	3.0	3.7	4.1	3.7
rubidus	14.2	12.3	8.6	4.4	4.3	4.3	5.0	3.6	3.9	3.1	3.5	3.2	2.8	1.8	1.7	1.8	3.3	-	1.9	2.2	1.4	1.7	2.8	0.9	1.0	1.9
ogilbyi	14.1	11.1	8.6	4.6	4.5	4.5	4.9	3.8	3.2	2.9	2.8	2.8	1.8	2.5	0.9	0.7	1.7	2.0	-	2.2	0.9	3.4	2.7	3.0	3.1	3.6
niger	14.0	12.4	8.9	5.0	4.9	5.2	5.3	3.9	3.9	2.8	3.2	3.5	2.5	2.9	1.8	2.0	3.8	2.2	2.2	-	1.5	3.2	2.8	2.8	3.0	3.2
weynsii	14.1	11.7	8.5	3.9	3.7	3.7	4.2	3.3	3.1	2.8	3.2	2.6	2.2	2.1	0.5	0.6	2.4	1.4	0.9	1.5	-	2.7	2.3	2.3	2.4	3.0
rufilatus	14.4	13.1	8.6	5.2	5.0	5.0	5.7	4.6	4.9	4.0	4.5	3.9	4.3	2.6	3.0	3.2	3.6	1.7	3.5	3.3	2.8	-	3.5	1.5	1.7	0.5
adersi	14.1	12.3	7.8	4.3	4.5	4.2	4.6	3.9	3.9	3.9	4.3	3.8	3.9	3.2	2.4	2.5	3.1	2.9	2.8	2.9	2.4	3.6	-	3.6	4.0	3.7
natalensis	14.6	13.1	8.4	5.3	5.2	5.2	5.9	4.5	4.7	3.3	4.0	3.7	3.6	2.4	2.6	2.8	3.9	0.9	3.0	2.9	2.4	1.5	3.7	-	0.3	1.3
harveyi	15.0	13.1	8.6	5.7	5.6	5.6	5.7	4.3	4.6	3.7	4.2	3.9	3.5	2.8	2.8	2.9	4.3	1.0	3.2	3.0	2.5	1.7	4.2	0.3	-	1.4
nigrifrons	14.1	13.1	8.3	5.4	5.3	5.3	6.0	4.6	4.9	4.0	4.5	3.9	4.3	2.9	3.3	3.5	3.9	2.0	3.7	3.3	3.0	0.5	3.9	1.3	1.4	-

Rate heterogeneity

The relative rate test (Robinson *et al.* 1998) identified a single species, *C. monticola*, as having a different substitution rate relative to its congeners. Four pairwise comparisons involving the three *C. monticola* lineages resulted in significant P values for synonymous changes. These were *C. monticola* vs.: *C. callipygus* (P=0.005); *C. weynsi* (P=0.004); *C. harveyi* (P=0.005); *C. adersi* (P=0.002). Since no rate variation was detected between the three *C. monticola* specimens analyzed it would appear that the lineage itself is characterized by a different rate. When excluding transitions, and considering transversional changes in isolation, none of the 19 duiker lineages are characterized by significant rate heterogeneity. In contrast to the cytochrome *b* gene, analysis of the small ribosomal sequences revealed only a single case of rate heterogeneity involving *C. natalensis* and *C. ogilbyi* (P=0.006). This is not a novel finding since different rates of evolution between cytochrome *b* and the small ribosomal subunit have been reported (Zhang & Ryder 1995, Matthee & Robinson 1997), perhaps reflecting the different evolutionary constraints under which these loci have evolved. As was the case with cytochrome *b*, no rate heterogeneity was detected when analyses were based on transversional changes alone.

Cytochrome *b*

The cytochrome *b* data set consisted of 1 140 bp of which 311 characters were variable and 243 were parsimony informative for the ingroup. As is typical for protein coding genes, most of these variable sites were at the third codon position (259 variable sites, 207 parsimony informative sites), with the second codon position being the most conserved (14 variable sites, 8 parsimony informative sites). First codon positions had 38 variable and 28 parsimony informative characters, six of which comprised leucine changes. Synonymous changes far outnumbered non-synonymous changes, resulting in only 28 variable and 18 parsimony informative amino acids for the ingroup.

The nucleotide composition for cytochrome *b* is given in Table 6. As is typical for most mammalian lineages (Irwin *et al.* 1991), the duiker lineages display an overall bias against guanine (13.2%), with the selection against this nucleotide most pronounced at third codon positions (3.5% guanine). Transitional changes outnumbered transversions 6.3 to 1 when counting all changes across 23 ingroup and 3 outgroup taxa. This ratio increased to 13.2 to 1 when the outgroups were excluded, and when considering a single representative for each cephalophine species, the ratio was 13.7 to 1.

Table 6 Base composition for the complete cytochrome *b* gene, as well as first, second, and third codon positions. Base frequencies were calculated in Paup 4.0b2a (Swofford 1999). Abbreviations used to indicate sampling localities are: Cam=Cameroon; Gab=Gabon; SA=South Africa; CAR=Central African Republic; Zim=Zimbabwe; Con=Democratic Republic of Congo.

	Complete gene				First				Second				Third			
Taxon	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T
Antilocapra	0.325	0.298	0.128	0.247	0.302	0.260	0.221	0.215	0.200	0.239	0.136	0.423	0.473	0.394	0.028	0.102
Oreotragus	0.317	0.276	0.130	0.275	0.310	0.250	0.213	0.226	0.202	0.247	0.136	0.413	0.439	0.331	0.042	0.186
Cow	0.312	0.301	0.134	0.251	0.286	0.260	0.228	0.223	0.205	0.252	0.136	0.405	0.444	0.392	0.036	0.126
monticola(Cam)	0.314	0.293	0.127	0.264	0.294	0.255	0.231	0.218	0.205	0.247	0.134	0.413	0.444	0.378	0.015	0.160
monticola(Gab)	0.314	0.289	0.131	0.264	0.294	0.260	0.231	0.213	0.202	0.244	0.136	0.415	0.447	0.363	0.026	0.163
monticola(SA)	0.313	0.291	0.128	0.267	0.294	0.255	0.228	0.221	0.205	0.247	0.134	0.413	0.439	0.371	0.021	0.168
maxwellii	0.311	0.285	0.133	0.269	0.297	0.247	0.228	0.226	0.202	0.247	0.136	0.413	0.434	0.363	0.034	0.168
grimmia(CAR)	0.314	0.300	0.128	0.256	0.297	0.271	0.223	0.207	0.205	0.252	0.136	0.405	0.439	0.378	0.026	0.155
grimmia(Zim)	0.317	0.298	0.126	0.257	0.297	0.271	0.223	0.207	0.210	0.252	0.134	0.402	0.444	0.371	0.021	0.163
silvicultor	0.313	0.298	0.128	0.260	0.297	0.257	0.226	0.218	0.207	0.244	0.134	0.413	0.434	0.392	0.023	0.150
spadix	0.313	0.298	0.129	0.258	0.302	0.260	0.223	0.213	0.207	0.250	0.134	0.407	0.428	0.384	0.031	0.155
dorsalis	0.314	0.296	0.127	0.262	0.302	0.257	0.221	0.218	0.207	0.252	0.134	0.405	0.431	0.378	0.026	0.163
jentinki	0.314	0.284	0.128	0.272	0.302	0.257	0.223	0.215	0.200	0.250	0.142	0.407	0.439	0.344	0.021	0.194
leucogaster	0.309	0.299	0.135	0.256	0.297	0.255	0.226	0.221	0.205	0.250	0.136	0.407	0.426	0.392	0.042	0.139
callipygus(Gab)	0.306	0.295	0.137	0.260	0.300	0.257	0.223	0.218	0.205	0.247	0.136	0.410	0.413	0.381	0.052	0.152
callipygus(Con)	0.306	0.297	0.137	0.258	0.300	0.257	0.223	0.218	0.205	0.247	0.136	0.410	0.413	0.386	0.052	0.147
zebra	0.313	0.305	0.129	0.249	0.300	0.263	0.223	0.213	0.202	0.250	0.139	0.407	0.436	0.402	0.026	0.126
rubidus	0.318	0.295	0.126	0.258	0.294	0.260	0.228	0.215	0.205	0.247	0.136	0.410	0.456	0.377	0.013	0.147
ogilbyi	0.310	0.304	0.131	0.253	0.292	0.273	0.228	0.205	0.205	0.247	0.136	0.410	0.434	0.392	0.028	0.144
niger	0.314	0.295	0.130	0.259	0.300	0.260	0.223	0.215	0.205	0.247	0.136	0.410	0.436	0.378	0.031	0.152
weynsi	0.306	0.289	0.137	0.266	0.300	0.255	0.223	0.221	0.205	0.247	0.136	0.410	0.413	0.365	0.052	0.168
rufilatus	0.302	0.305	0.140	0.251	0.297	0.257	0.223	0.221	0.205	0.250	0.136	0.407	0.405	0.407	0.060	0.126
adersi	0.309	0.295	0.133	0.260	0.297	0.257	0.226	0.218	0.205	0.252	0.136	0.405	0.426	0.376	0.036	0.157
natalensis	0.307	0.300	0.137	0.253	0.300	0.252	0.223	0.223	0.207	0.250	0.136	0.405	0.415	0.400	0.052	0.131
harveyi	0.304	0.298	0.141	0.256	0.300	0.252	0.223	0.223	0.207	0.250	0.136	0.405	0.405	0.392	0.063	0.139
nigrifrons	0.306	0.307	0.137	0.249	0.300	0.252	0.223	0.223	0.205	0.250	0.136	0.407	0.413	0.418	0.052	0.115
Mean	0.311	0.296	0.132	0.259	0.298	0.258	0.225	0.217	0.205	0.248	0.136	0.409	0.432	0.381	0.035	0.150

The evaluation of 10 000 random trees showed that the data contained significantly more signal than expected at random (26 taxa: g1=-0.699; P<0.01; Hillis & Huelsenbeck 1992). Since phylogenetic signal is not necessarily evenly distributed throughout the topology (Hillis 1991, Hillis & Huelsenbeck 1992), g1 values were recalculated for subsets of the data. The g1 values were significant even when various combinations of taxa, including species at the well supported tips, were excluded.

Cytochrome *b* phylogeny

Using all three outgroups, and the empirical ingroup ti/tv ratio (13:1) and heuristic search option in Paup 4.0b2a, parsimony analyses of the 22 taxa (three outgroup taxa and a single representative of each duiker species) resulted in a single tree of 2739 steps (Fig. 2). Four duiker clades were retrieved: the conservative dwarf lineage *(C. monticola, C. maxwellii)*, the savanna specialist *(S. grimmia)*, the giant duiker group *(C. silvicultor, C. spadix, C. dorsalis, C. jentinki)* and the red duiker lineage. Within the latter, two distinct maternal clusters are evident: an east African red duiker clade (sensu Kingdon 1982) comprising *C. leucogaster, C. rufilatus, C. nigrifrons, C. natalensis, C. harveyi* and a west African red duiker clade (sensu Kingdon 1982) comprising *C. callipygus, C. weynsi, C. ogilbyi, C. rubidus,* and *C. niger*. There is good agreement between these mitochondrially defined groups and those which have previously been delimited on phenotypic grounds, principally the length and texture of hair on the shoulders (Kingdon 1982). The only exceptions are *C. leucogaster* which Kingdon (1982) places in the west African group, and *C. rubidus* which is allocated to the east African group. Analysis of the data using a 2:1 and unordered weighting scheme resulted in several equally most parsimonious solutions (three and eight trees respectively); however, the consensus trees (not shown) were near identical to the empirical tree, differing only with respect to associations that received less than 50% bootstrap support. Both of the consensus trees indicate a sister group relationship between *C. adersi* and *C. niger*, with *C. zebra*, the red duiker lineage, and the giant duiker forming a trichotomy.

Two trees, one step longer than the most parsimonious solution (Fig. 2), were recovered. The consensus of these differed from Figure 2 in the positioning of *C. zebra* and *C. adersi* (which clustered as sister taxa basal to the red duiker lineage as opposed to being basal within the forest duiker), and in the unresolved placement of *C. niger* within the red duiker lineage. However, nonparametric bootstrap values for nodes supporting *C. zebra*, *C. adersi*, and *C. niger* were invariably low (<50%) and the resolution of these species in the

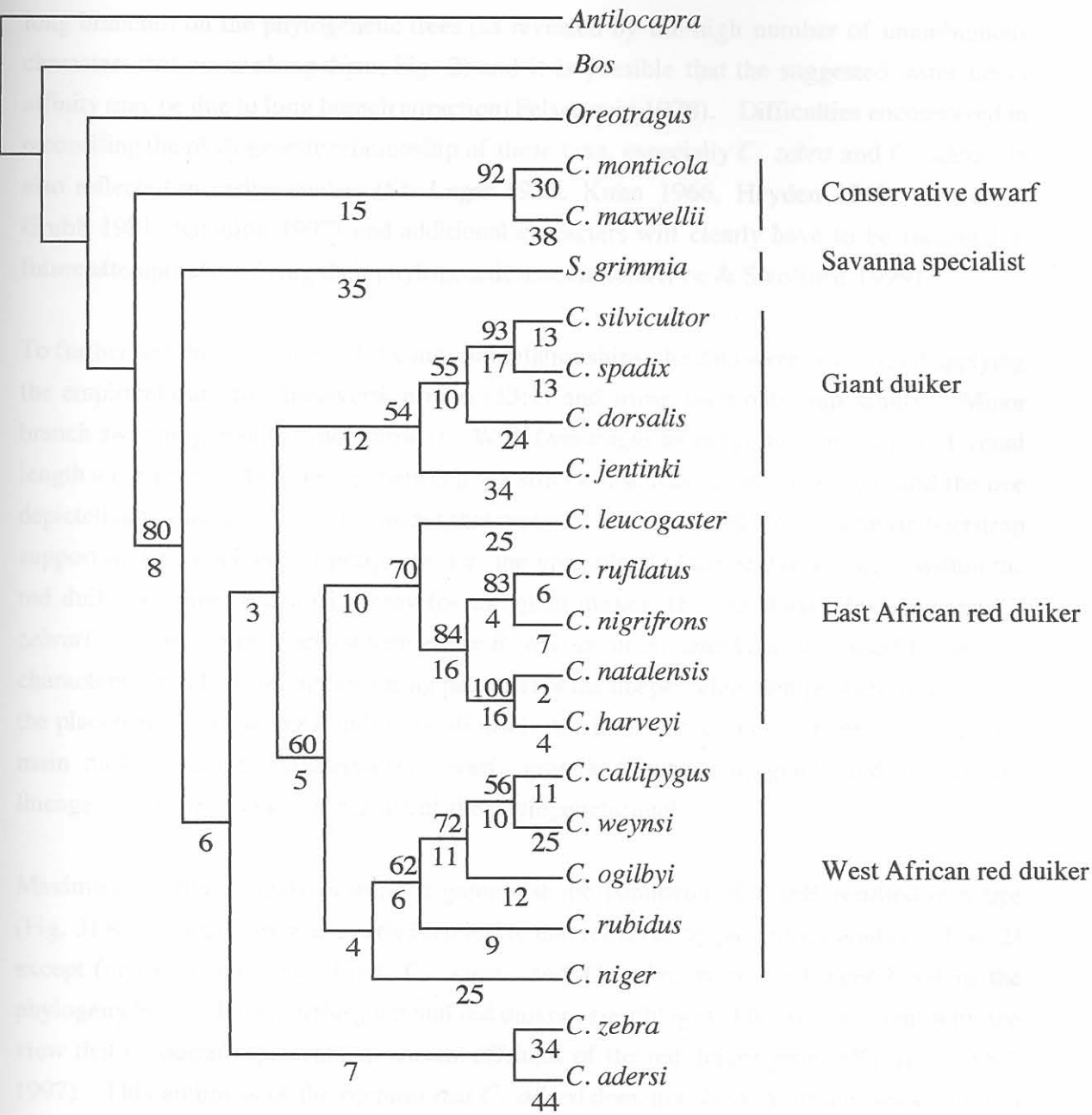


Figure 2 The most parsimonious tree obtained under the empirical cytochrome *b* ti/tv weighting of 13:1. This tree is based on 1 140 bp and has a length of 2 739 steps. Nonparametric bootstrap values obtained after 1 000 iterations are shown above branches while the number of unambiguous characters that occur along branches are shown below them. Bootstrap values for nodes that received less than 50% support are not indicated.

duiker phylogeny unfortunately remains obscure irrespective method of analyses or weighting scheme used. Moreover, *C. zebra* and *C. adersi* are characterized by relatively long branches on the phylogenetic trees (as revealed by the high number of unambiguous characters that occur along them; Fig. 2) and it is possible that the suggested sister taxon affinity may be due to long branch attraction (Felsenstein 1978). Difficulties encountered in reconciling the phylogenetic relationship of these taxa, especially *C. zebra* and *C. adersi*, is also reflected in earlier studies (St. Leger 1936, Kuhn 1966, Heyden 1968, Groves & Grubb 1981, Kingdon 1997) and additional characters will clearly have to be included in future attempts at resolving their phylogenetic associations (Poe & Swofford 1999).

To further test the robustness of the ingroup relationships, the data were reanalyzed applying the empirical transition/transversion ratio (13:1) and using each outgroup singly. Minor branch swapping resulted (not shown). With *Oreotragus* as outgroup, three trees of equal length were found. Differences between the strict consensus of the three trees and the tree depicted in Figure 2 involved the nodes that received less than 50% nonparametric bootstrap support in the combined outgroup tree, i.e. the unresolved placement of *C. niger* within the red duiker lineage, and a tricotomy for the giant duiker, the red duiker lineage, and *C. zebra*/*C. adersi*. In instances where either *B. taurus* or *A. americana* was used to polarize characters, slightly different branching patterns for the deeper relationships were found (e.g. the placement of *Sylvicapra* and its relationship to the forest duiker). However, the four main duiker lineages (conservative dwarf, grassland specialist, giant, and red duiker lineages) remained a constant feature of the phylogenetic analyses.

Maximum likelihood analysis using a gamma shape parameter of 0.168 resulted in a tree (Fig. 3) whose topology was nearly identical to that retrieved by parsimony analysis (Fig. 2) except for two disparities. First, C. *adersi* and *C. zebra* were no longer basal in the phylogeny but fell between the giant and red duiker assemblages. This is consistent with the view that *C. adersi* represents an ancient offshoot of the red duiker group (Kingdon 1982, 1997). This author is of the opinion that *C. adersi* does not show a strong association to either of the two red duiker clades (east or west African) but is rather central to their radiation.

Secondly, contrary to the parsimony result (although admittedly with no bootstrap support) *C. adersi* and *C. zebra* are not grouped as sister taxa by maximum likelihood. Since this analysis does not suffer from the effects of long branch attraction (Huelsenbeck 1995, Swofford *et al.* 1996) their association by parsimony may be misleading. The tree generated

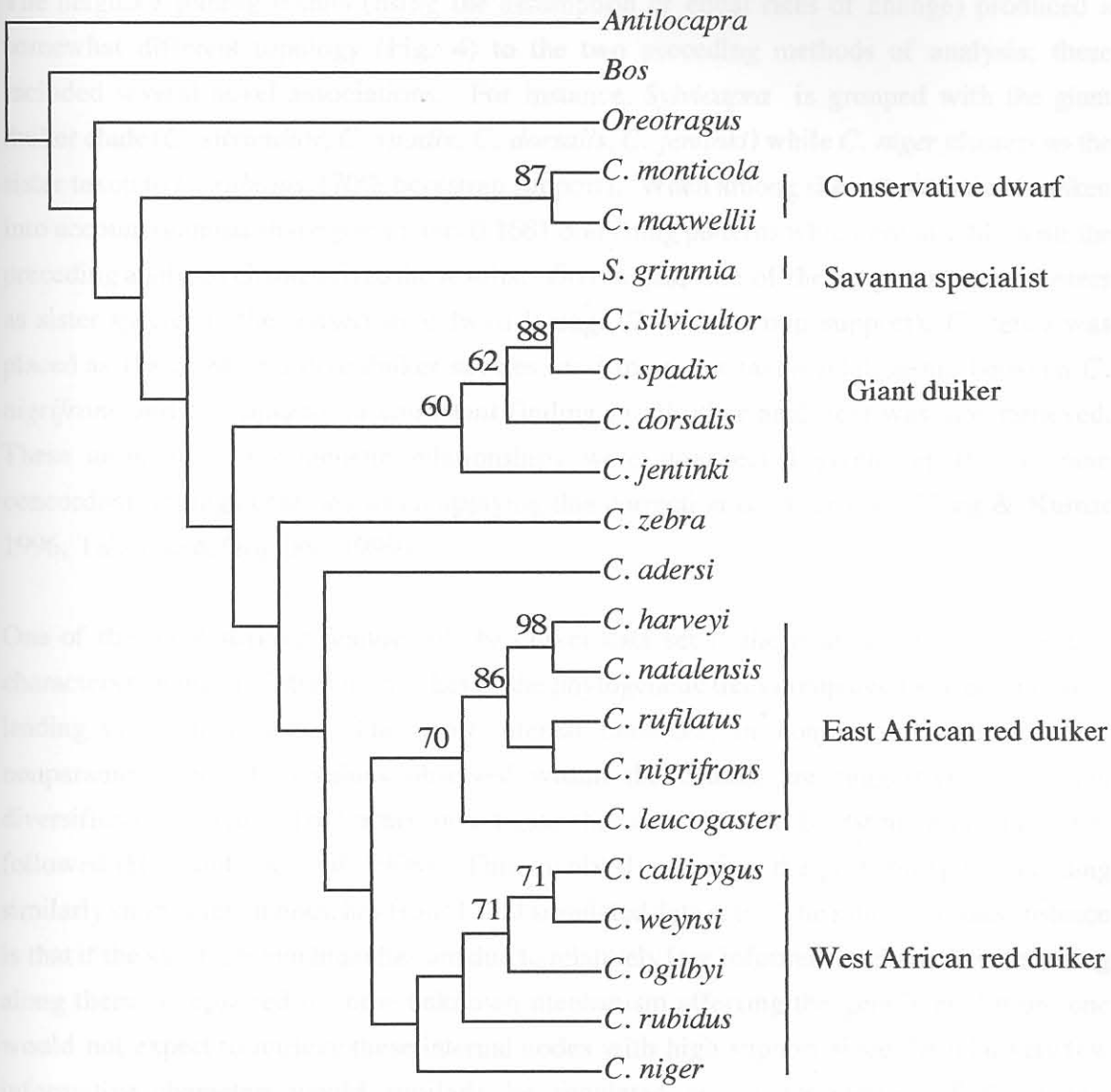


Figure 3 The optimal maximum likelihood topology inferred from the cytochrome *b* data. The tree was constructed in Paup 4.0b2a (Swofford 1999) under the HKY (Hasegawa *et al.* 1985) model of sequence evolution with rate heterogeneity (gamma shape parameter=0.168), empirical ti/tv ratio (13:1), and nucleotide frequencies calculated from the data. Bootstrap support, obtained from 100 iterations, is shown.

by maximum likelihood analysis under the assumption of equal rates of change (not shown) was identical to that presented in Figure 3 with the exception of a sister taxon relationship for *C. adersi* and *C. zebra* which received less than 50% bootstrap support.

The neighbor joining results (using the assumption of equal rates of change) produced a somewhat different topology (Fig. 4) to the two preceding methods of analysis; these included several novel associations. For instance, *Sylvicapra* is grouped with the giant duiker clade *(C. silvicultor, C. spadix, C. dorsalis, C. jentinki)* while *C. niger* clusters as the sister taxon to *C. rubidus* (70% bootstrap support). When among site rate variation is taken into account (gamma shape parameter=0.168) branching patterns which are at odds with the preceding analyses characterized the results. *Oreotragus*, one of the outgroup taxa, clusters as sister species to the conservative dwarf lineage (74% bootstrap support), *C. zebra* was placed as the most primitive duiker species, and the sister taxon relationship between *C. nigrifrons* and *C. rufilatus* (a consistent finding in all other analyses) was not retrieved. These anomalous phylogenetic relationships were unexpected given reports of more concordant findings obtained when applying this correction to other taxa (Yang & Kumar 1996, Takezaki & Gojobori 1999).

One of the most striking features of the duiker data set is the relatively few informative characters that map to internal branches of the phylogenetic trees compared to larger numbers leading to terminal taxa. The short internal branches, in conjunction with the low nonparametric bootstrap values observed within the clades, are suggestive of a rapid diversification event. To further investigate this a parametric bootstrap procedure was followed (Huelsenbeck *et al.* 1996). This involved assessing the probability of obtaining similarly short internal branches from 1 000 simulated data sets. The rational in this instance is that if the short internal branches are due to relatively few informative characters occurring along them, as opposed to some unknown mechanism affecting the gene's evolution, one would not expect to retrieve these internal nodes with high support since the relatively few informative characters would similarly be simulated as an assumption of the model (Halanych & Robinson 1999). Importantly for the present investigation, this approach led to a parametric bootstrap topology which is identical to the cytochrome *b* parsimony topology. High bootstrap support (Fig. 5) is found for the main duiker lineages (conservative dwarf, giant duiker, and red duiker lineage). In sharp contrast, the internal branches received low support, indicating the relatively few characters defining earlier speciation events within the group.

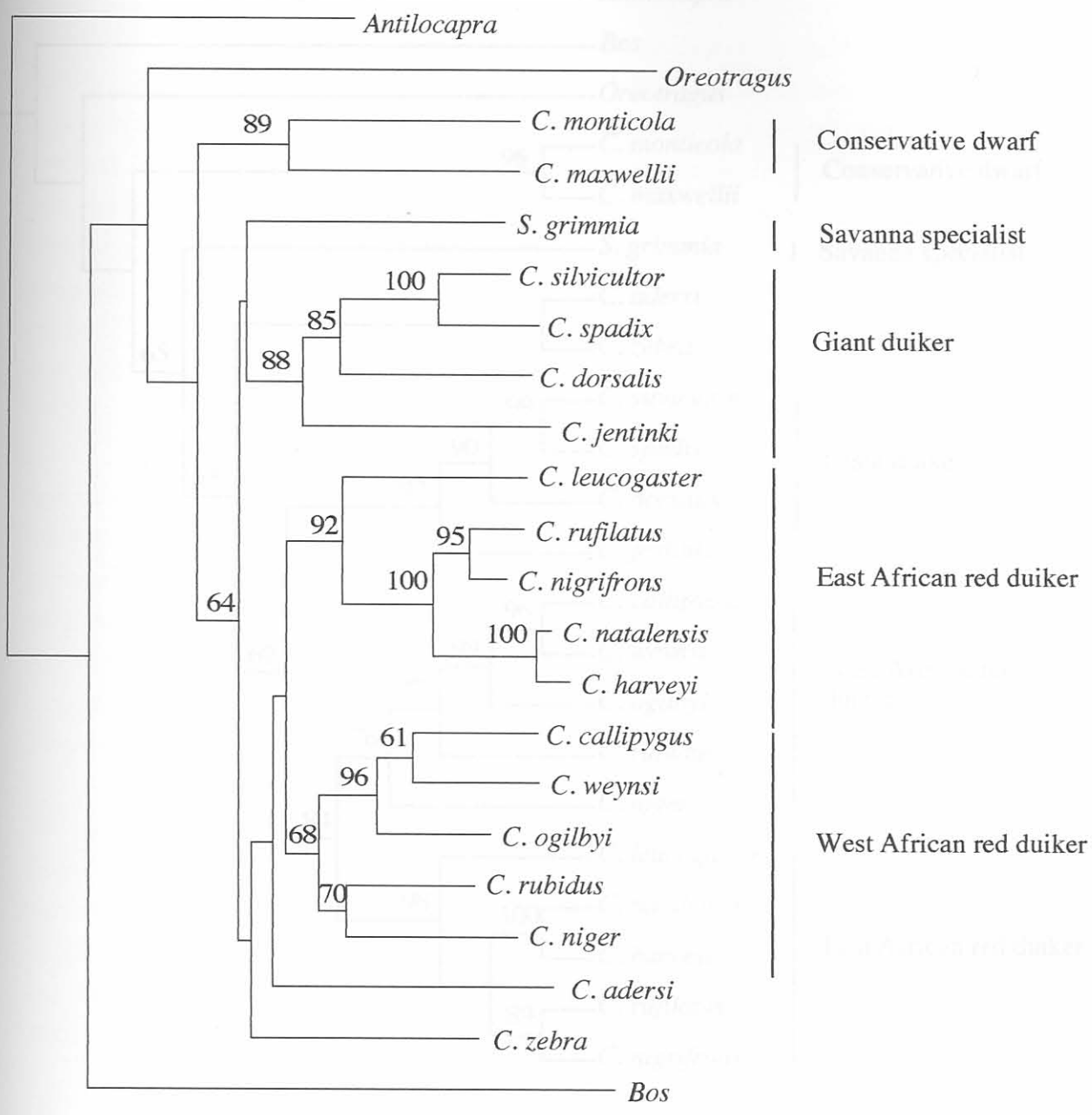


Figure 4 Neighbor joining tree constructed in Paup 4.0b2a (Swofford 1999) from 1 140 bp of the cytochrome *b* gene. The HKY (Hasegawa *et al.* 1985) correction was applied to the distance matrix, with equal rates of evolution assumed for all lineages. The branch lengths are drawn proportionally to the amount of change occurring along them, and nonparametric bootstrap support is indicated.

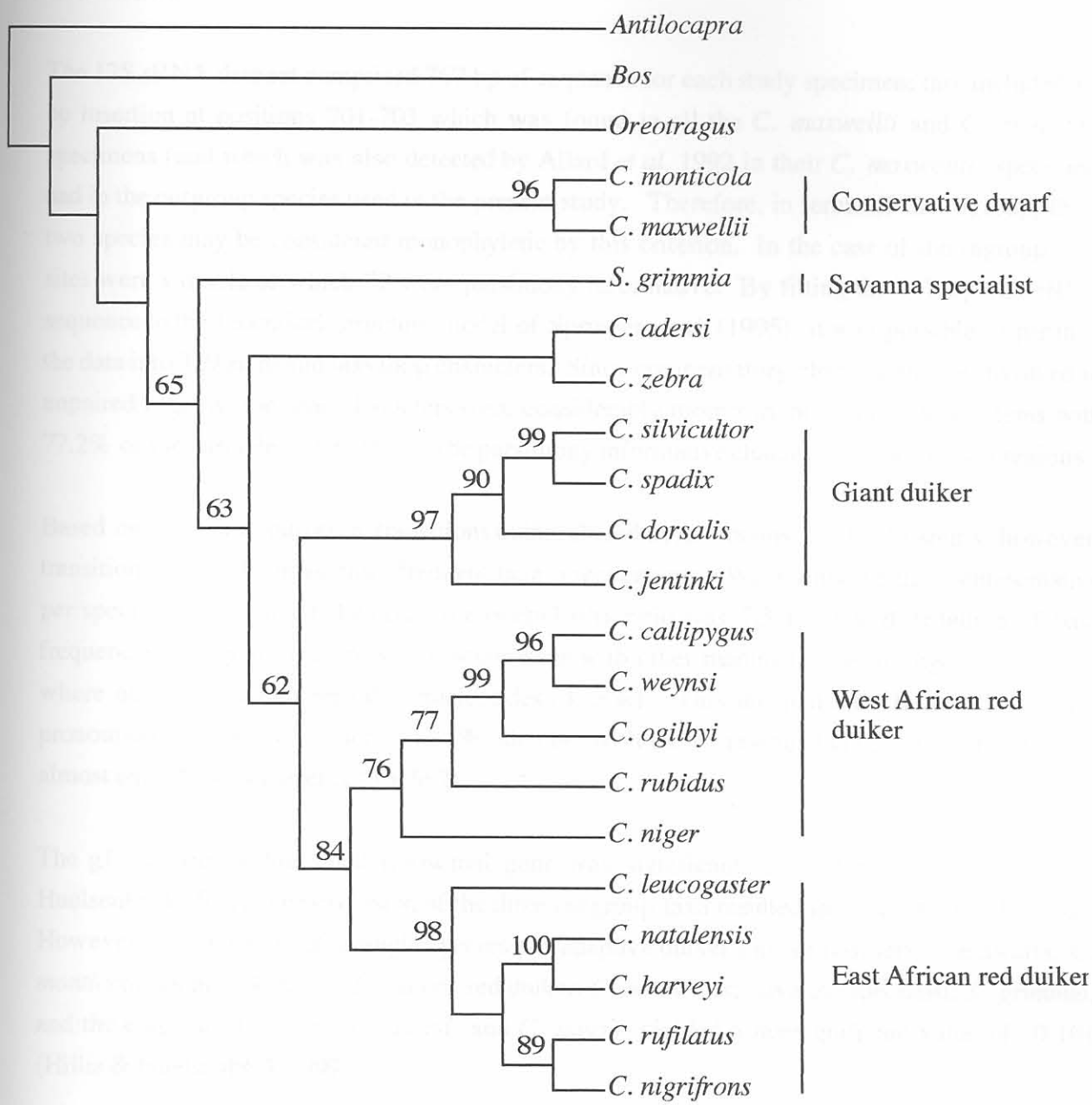


Figure 5 The cytochrome *b* parametric bootstrap topology obtained from 1 000 simulated data sets created by the Siminator computer program (Huelsenbeck *et al.* 1996). The most parsimonious tree (Fig. 2) was used as the model tree, with the gamma shape parameter (0.168) and kappa value (27.167) estimated *via* maximum likelihood. The simulated data sets were developed under the HKY (Hasegawa *et al.* 1985) model of evolution.

12S rRNA

The 12S rRNA data set comprised 767 bp of sequence for each study specimen; this included a 3 bp insertion at positions 701-703 which was found in all the *C. maxwellii* and *C. monticola* specimens (and which was also detected by Allard *et al.* 1992 in their *C. maxwellii* specimen) and in the outgroup species used in the present study. Therefore, in terms of the ingroup, these two species may be considered monophyletic by this criterion. In the case of the ingroup, 110 sites were variable of which 72 were parsimony informative. By fitting the 767 bp 12S rRNA sequence to the secondary structure model of Springer *et al.* (1995), it was possible to partition the data into 359 stem and 408 loop characters. Since compensatory changes are not involved in unpaired regions, the loop characters were considerably more variable than those of stems with 77.2% of the variable and 81.9% of the parsimony informative characters found in loop regions.

Based on the 23 ingroup taxa, transitions outnumbered transversions 7.1:1. In stems, however, transitions were 9.8 times more frequent than transversions. When only a single representative per species was included (19 taxa), the overall ti/tv ratio was 7.3:1. The distribution of base frequencies across all characters is in accordance with other mammals (see Springer *et al.* 1995) where adenines outnumber other nucleotides (37.8%). This unequal base frequency is most pronounced in unpaired regions (43.1% adenine) with paired regions being characterized by an almost equal base occurrence (Table 7).

The g1 statistic for the small ribosomal gene was significant (g1=-0.853; P<0.01; Hillis & Huelsenbeck 1992). The exclusion of the three outgroup taxa resulted in a g1=-1.019 (P<0.01). However, the inclusion of a single species per adaptive duiker lineage (conservative dwarfs: *C. monticola*, giant duiker: *C. silvicultor*, red duiker: *C. natalensis*; savanna specialist: *S. grimmia*, and the enigmatic *C. zebra*, *C. adersi*, and *C. niger)* yielded a nonsignificant value of -0.101 (Hillis & Huelsenbeck 1992).

12S rRNA phylogeny

Parsimony analysis of the 12S rRNA sequences with a 2:1 weighting resulted in 14 equally most parsimonious trees of 386 steps. The consensus tree is presented in Figure 6. Moreover, when weakly supported nodes (<50%) were collapsed, the resultant topology was identical to the 12S rRNA bootstrap trees obtained under an unordered and empirical 7:1 weighting scheme. Treating the 3 bp indel as missing data, or as a fifth character, did not influence the result although in the latter instance marginally higher bootstrap support for the basal placement of

Table 7 Base composition for the small ribosomal subunit. The composition of the stem and loop domains are also indicated. Frequencies were calculated in Paup 4.0b2a (Swofford 1999). Abbreviations used to indicate sampling localities are: Cam=Cameroon; Gab=Gabon; SA=South Africa; CAR=Central African Republic; Zim=Zimbabwe; Con=Democratic Republic of Congo.

	Complete gene				Stem				Loop			
Taxon	A	C	G	T	A	C	G	T	A	C	G	T
Antilocapra	0.364	0.226	0.185	0.223	0.298	0.224	0.246	0.230	0.422	0.232	0.127	0.218
Oreotragus	0.371	0.228	0.177	0.222	0.310	0.229	0.229	0.229	0.423	0.235	0.127	0.213
Bos	0.381	0.222	0.176	0.219	0.308	0.227	0.235	0.229	0.444	0.223	0.120	0.211
monticola(Cam)	0.379	0.221	0.169	0.229	0.318	0.221	0.224	0.235	0.428	0.227	0.119	0.224
monticola(Gab)	0.378	0.224	0.170	0.226	0.318	0.221	0.224	0.235	0.425	0.232	0.122	0.220
monticola(SA)	0.382	0.221	0.165	0.229	0.317	0.227	0.224	0.230	0.437	0.224	0.110	0.227
maxwellii	0.378	0.227	0.171	0.223	0.314	0.227	0.227	0.230	0.433	0.232	0.117	0.217
grimmia(CAR)	0.379	0.226	0.174	0.219	0.321	0.221	0.221	0.235	0.428	0.238	0.127	0.204
grimmia(Zim)	0.379	0.226	0.175	0.218	0.321	0.218	0.221	0.237	0.428	0.240	0.130	0.200
silvicultor	0.379	0.226	0.172	0.221	0.313	0.224	0.229	0.232	0.436	0.236	0.118	0.209
spadix	0.373	0.226	0.179	0.221	0.313	0.224	0.229	0.232	0.421	0.236	0.132	0.209
dorsalis	0.380	0.231	0.172	0.214	0.318	0.227	0.224	0.229	0.433	0.243	0.122	0.200
jentinki	0.379	0.223	0.172	0.223	0.316	0.221	0.227	0.235	0.433	0.233	0.120	0.212
leucogaster	0.380	0.225	0.171	0.222	0.313	0.227	0.227	0.232	0.438	0.231	0.118	0.212
callipygus(Gab)	0.380	0.229	0.172	0.217	0.321	0.221	0.221	0.235	0.431	0.243	0.125	0.200
callipygus(Con)	0.379	0.229	0.174	0.217	0.318	0.221	0.224	0.235	0.431	0.243	0.125	0.200
zebra	0.380	0.221	0.171	0.226	0.313	0.218	0.227	0.240	0.438	0.231	0.118	0.212
rubidus	0.377	0.226	0.173	0.222	0.316	0.221	0.224	0.237	0.429	0.236	0.123	0.210
ogilbyi	0.380	0.225	0.172	0.221	0.318	0.221	0.224	0.235	0.433	0.236	0.122	0.207
niger	0.382	0.222	0.170	0.225	0.321	0.224	0.221	0.232	0.433	0.228	0.120	0.216
weynsi	0.382	0.227	0.171	0.218	0.321	0.221	0.221	0.235	0.433	0.240	0.122	0.202
rufilatus	0.378	0.227	0.170	0.223	0.313	0.221	0.227	0.237	0.433	0.238	0.115	0.212
adersi	0.384	0.221	0.170	0.223	0.316	0.221	0.227	0.235	0.443	0.228	0.115	0.212
natalensis	0.374	0.230	0.174	0.221	0.313	0.221	0.227	0.237	0.426	0.243	0.122	0.207
harveyi	0.371	0.229	0.176	0.222	0.313	0.221	0.227	0.237	0.421	0.240	0.127	0.209
nigrifrons	0.376	0.227	0.171	0.223	0.310	0.221	0.229	0.237	0.433	0.238	0.115	0.212
Mean	0.378	0.226	0.173	0.222	0.315	0.223	0.226	0.234	0.431	0.235	0.121	0.210

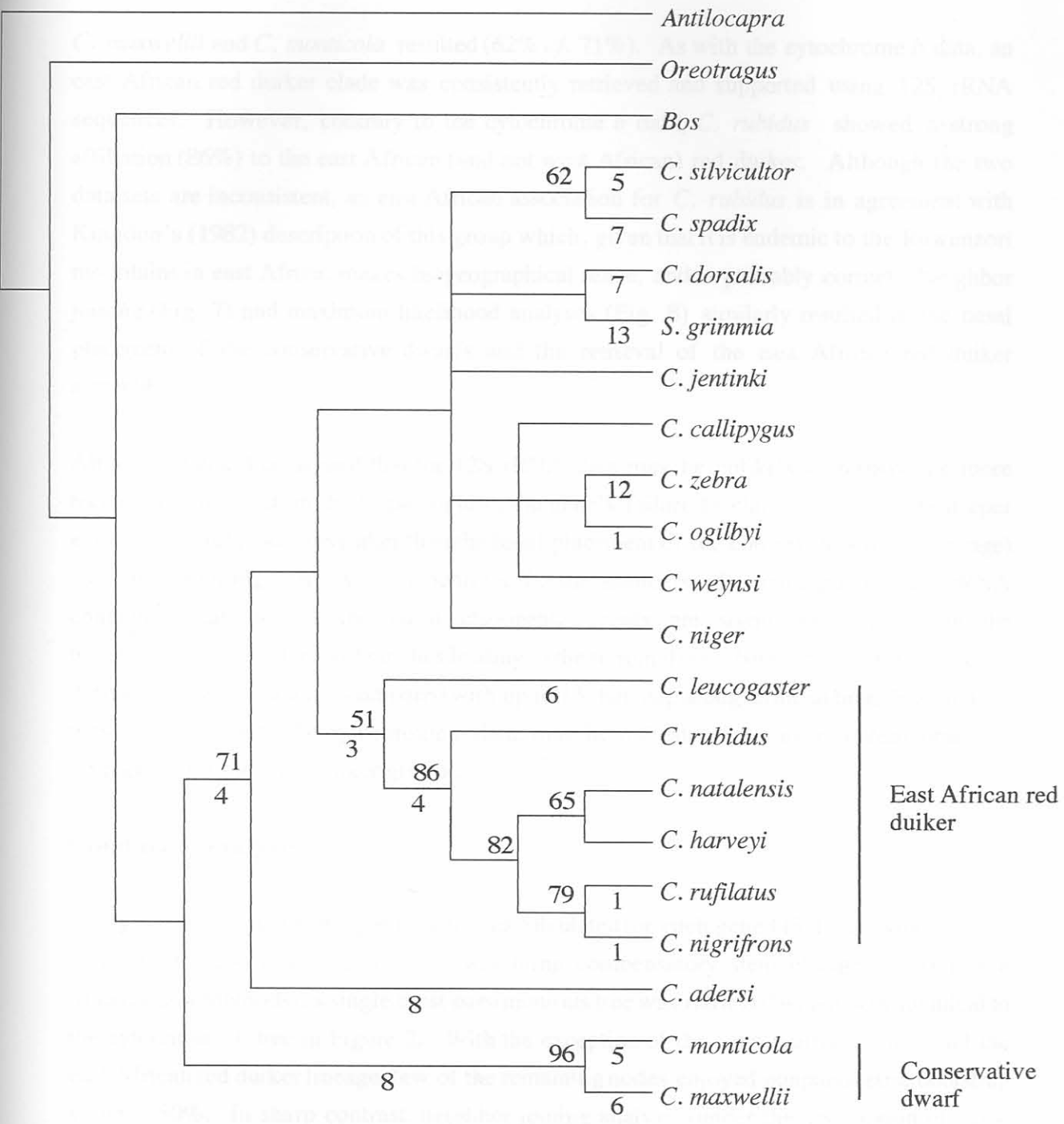


Figure 6 Strict consensus tree of 14 equally most parsimonious trees obtained from 767 bp of the 12S rRNA gene under a 2:1 ti/tv ratio. Nonparametric bootstrap values for nodes are presented above branches and the number of unambiguous changes that plot along them are indicated below the branches. No changes are shown on branches within polytomies.

C. maxwellii and *C. monticola* resulted (62% *cf.* 71%). As with the cytochrome *b* data, an east African red duiker clade was consistently retrieved and supported using 12S rRNA sequences. However, contrary to the cytochrome *b* data, *C. rubidus* showed a strong affiliation (86%) to the east African (and not west African) red duiker. Although the two data sets are inconsistent, an east African association for *C. rubidus* is in agreement with Kingdon's (1982) description of this group which, given that it is endemic to the Ruwenzori mountains in east Africa, makes biogeographical sense, and is probably correct. Neighbor joining (Fig. 7) and maximum likelihood analyses (Fig. 8) similarly resulted in the basal placement of the conservative dwarfs and the retrieval of the east African red duiker assemblage.

Although it could be argued that the 12S rRNA data may be unlikely to resolve the more recent radiations within the Cephalophini, the gene's failure to clarify many of the deeper evolutionary relationships (other than the basal placement of the conservative dwarf lineage) was disappointing. As with cytochrome *b*, the numbers of unambiguous 12S rRNA characters that support the basal cladogenic events are significantly less than the unambiguous characters on branches leading to the terminal taxa (between 3 and 4 characters define the deeper branches compared with up to 13 that map along terminal branches) and, as was speculated with the cytochrome *b* data, may be indicative of a near contemporaneous divergence for the main duiker groups.

Combined analysis

Using the empirical ti/tv weighting schemes calculated for each gene (13:1 for cytochrome *b* and 7:1 for 12S rRNA), and down weighting compensatory stem changes by 0.6 (see Material and Methods), a single most parsimonious tree was retrieved which was identical to the cytochrome *b* tree in Figure 2. With the exception of the conservative dwarfs and the east African red duiker lineage, few of the remaining nodes enjoyed nonparametric bootstrap values >50%. In sharp contrast, neighbor joining analysis (under the assumption of equal rates) resulted in a well supported tree whose topology was identical to that presented in Figure 4. The maximum likelihood tree (estimated by applying equal rates of change and the empirical ti/tv ratio of 11:1 for the combined data set) was identical to that retrieved using parsimony (see Fig. 2).

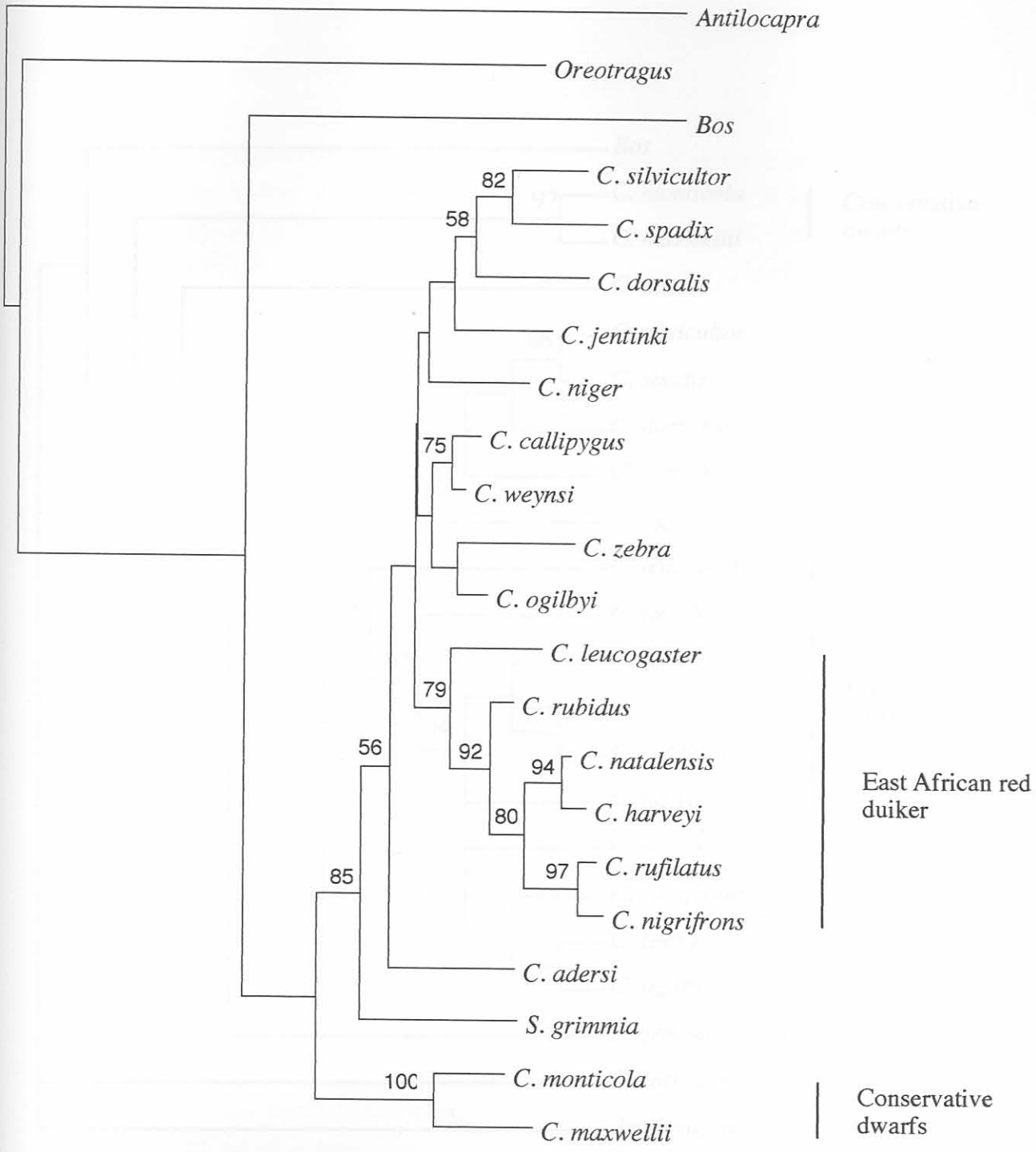


Figure 7 Neighbor joining tree constructed from 767 bp of the small ribosomal subunit in Paup 4.0b2a (Swofford 1999). The HKY correction (Hasegawa *et al.* 1985) was applied to the distance matrix. Branches are drawn proportionally to the amount of change occurring on them, and nonparametric bootstrap support for nodes is indicated.

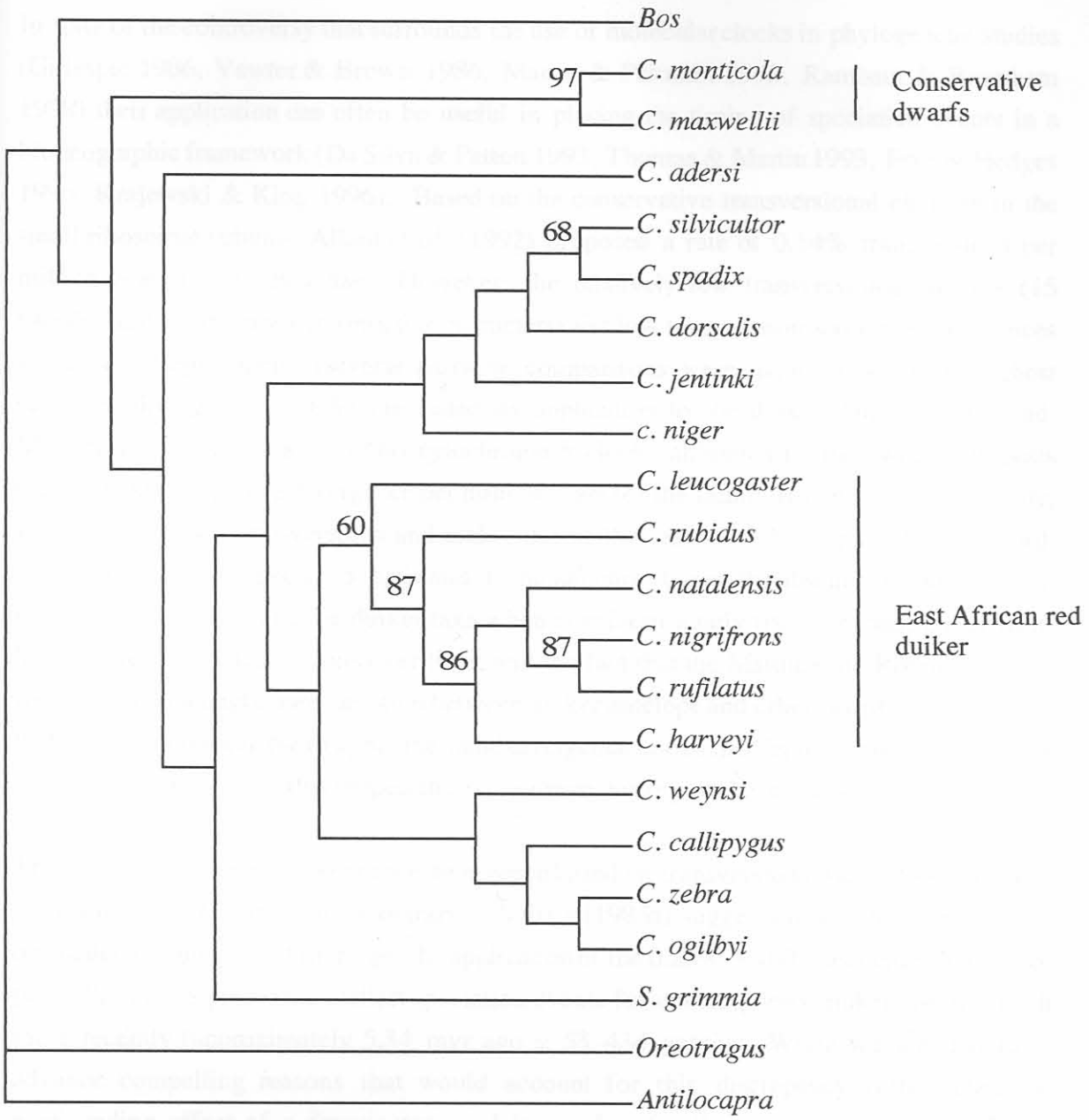


Figure 8 Maximum likelihood topology obtained from the 12S rRNA gene. The tree was constructed in Paup 4.0b2a (Swofford 1999) under the HKY (Hasegawa *et al.* 1985) model of sequence evolution with equal rates, the empirical ti/tv ratio (7:1), and nucleotide frequencies calculated from the data. Bootstrap support was obtained using 100 iterations.

Speciation events

In spite of the controversy that surrounds the use of molecular clocks in phylogenetic studies (Gillespie 1986, Vawter & Brown 1986, Martin & Palumbi 1993, Rambaut & Bromham 1998) their application can often be useful in placing the timing of speciation events in a biogeographic framework (Da Silva & Patton 1993, Thomas & Martin 1993, Frye & Hedges 1995, Krajewski & King 1996). Based on the conservative transversional changes in the small ribosomal subunit, Allard *et al.* (1992) proposed a rate of 0.14% transversions per million years for the Bovidae. However, the relatively few transversional changes (15 variable and 8 parsimony informative characters) and low transversion sequence divergences within the Cephalophini (several pairwise comparisons were identical, with the highest sequence divergence=1.18%) precluded its application to the duiker data set. Instead, Matthee and Robinson's (1999a) cytochrome *b* clock calibration is used which suggests 0.23% ± 0.01 sequence divergence per million years for the family Bovidae. This estimate, which is based on transversions and makes use of the extensive Alcelaphini fossil record, included two duiker species *(S. grimmia, C. monticola).* Given the absence of rate variation in the cytochrome *b* gene for duiker taxa when considering only transversions, the obvious linear increase of transversions over time, and the fact that the Matthee and Robinson study showed no meaningful rate variation between duiker antelope and other bovid taxa, suggests that it is not unreasonable to apply the same divergence rate to the Cephalophini in an attempt to explain the processes that shaped the evolutionary history of these antelope.

The highest cytochrome *b* sequence divergence based on transversions was 1.23% (between *C. monticola* and *C. rubidus).* Contrary to Vrba's (1985a) suggestion that the Cephalophini originated roughly 12-14 myr ago, the application of the 0.23% ± 0.01 sequence divergence per million years places the earliest speciation events for contemporary duiker species much more recently (approximately 5.34 myr ago ± 53 434 years). While we are unable to advance compelling reasons that would account for this discrepancy (other than the confounding effect of a fragmentary and incomplete fossil record) it is striking that at roughly this time (5 myr ago) several other bovid tribes similarly make their first appearance in the fossil record (Vrba 1985a). Hamilton (1982) and Boaz (1985) have speculated that climatic fluctuations became more marked towards the end of the Miocene (5 myr ago) which, coupled to an increase in the ice sheet covering Antarctica and subsequent lowering in temperatures (Lovett 1993), brought about a more arid climate with a possible reduction of the forest biome.

In addition to the climatic changes at the end of the Miocene, rifting in east Africa caused the uplift of the central African plateau leading to the separation of the main Guineo-Congolian forest block from the east African forests and the development of an arid north-south corridor between them (Lovett 1993). It is generally agreed (Grubb 1978, Livingstone 1982) that in contrast to the more stable west African environment, no single large refugium could have existed in east Africa but rather that the region comprised an archipelago of montane and lowland/coastal forests, with volcanic activity and rainshadows further contributing to the instability of forest habitat (Livingstone 1982, Wasser & Lovett 1993). If duiker speciation has been driven predominantly by habitat fragmentation and the concomitant disruption of gene flow between demes, one may anticipate that this would be reflected in the occurrence of taxa. Indeed, the giant duiker lineage (with its west African distribution) comprises only four species compared to the diverse red duiker lineage which includes 11 of the 18 forest duiker species. Although clearly speculative, one could also suggest that the species inhabiting the more stable west African forests should be of older age compared to east African species, where forest fragmentation during the more recent past (Moreau 1963, Livingstone 1975, Cooke 1978, Hamilton 1982) would have presented ample opportunities for allopatric speciation. This biogeographic scenario is well supported by the mitochondrial data in which the oldest Cephalophini lineages are those with a west African distribution. The average cytochrome *b* transversion sequence divergence within the almost exclusively west African giant duiker clade would suggest that the group is approximately 2.03 million years old, compared to both the west-central (sensu Kingdon 1982) and east African (sensu Kingdon 1982) red duiker species groups which are far more recent (985 506 and 695 625 years old respectively).

There has been much debate concerning the abilities of forest adapted species to adjust to life in an open savanna habitat. Dubost (1968) has argued that forest habitats may enforce such a strong and specialized adaptation on taxa that it significantly impedes the colonization of other ecotypes by them. This view is also held by Kingdon (1982) who posits that *Sylvicapra* has always been a savanna species which never entered the African forests. In sharp contrast, however, Grubb (1978) believes (based mainly on the derived nature of savanna species compared to their forest dwelling conspecifics and congeners) that the transition between habitats has primarily been from forest to savanna. This view is consistent with the mitochondrial DNA data which show *Sylvicapra* to be more derived than the conservative dwarf lineage *(C. monticola, C. maxwellii).* However, the divergence of *Sylvicapra* from the forest clade occurred early in the group's evolutionary history, since it predated the appearance of the other forest dwelling taxa.

Taxonomic implications

The results of this investigation question, and in some instances support, several of the nomenclatural divisions in current use in the Cephalophini taxonomy but which have traditionally been a source of contention.

Recognition of the genus Philantomba

The recognition of *Philantomba* as valid genus name for the two dwarf species has been hotly debated (Hard 1969, Groves & Grubb 1981, Smithers 1983, Robinson *et al.* 1996, Kingdon 1997). One of the most striking and strongly supported phylogenetic associations to emerge from the mitochondrial DNA analysis is that both *C. monticola* and *C. maxwellii* are basal within the tribe, a finding which is consistent with their retention of several primitive morphological (constricted rostrum, lack of inguinal glands) and phenotypic characters (small size; Grubb 1978, Kingdon 1982). Moreover, sequence divergence values separating the conservative dwarfs from the rest of the forest taxa are of similar magnitude to those between *Sylvicapra* and *Cephalophus* which, when taken together with the morphological (Grubb 1978) and cytogenetic evidence (Robinson *et al.* 1996), would argue for the reinstatement of *Philantomba* as the valid genus name for *C. monticola* and *C. maxwellii*. The recognition of a separate generic status for these two taxa would secure *Cephalophus* monophyly.

Phylogenetic placement of C. zebra *and* C. adersi

The evolutionary relationships of two duiker species *(C. adersi, C. zebra)* remains problematic. Ader's duiker is confined entirely to the Island of Zanzibar and the Sokoke coastal forest in Kenya while *C. zebra*, one of the most distinct taxa within the forest group, has a restricted west African distribution (Sierra Leone, Liberia, Ivory Coast). Kingdon (1982) recognizes *C. adersi* as a primitive member of the red duiker lineage, not really closely allied to either the east or west African species groups. In the present study, this species showed a novel association, grouping for the most part as a sister taxon to *C. zebra.* However, whether this is a true reflection of their evolution history or merely an artifact of long branch attraction is moot. In turn, *C. zebra* has been described as a modified survivor of an early phase in the duiker evolution (Kingdon 1982) and has even been placed in its own genus or subgenus *(Cephalophula zebra*; Heyden 1968, Groves & Grubb 1981).

Mitochondrial DNA analyses fail to unequivocally place this taxon in any of the recognized groups, with different genes and methods of analyses either clustering the zebra duiker basally to the red duiker lineage or, alternatively, outside of the forest duiker clade altogether. Both genetic and morphological data would therefore seem to suggest an independent evolutionary trajectory for this species.

Subspecies status of C. harveyi

The lowest sequence divergence detected between any of the duiker species using either the cytochrome *b* or 12S rRNA data is between *C. natalensis* and *C. harveyi* (cytochrome *b*=0.7%; 12S rRNA=0.3%). These values are comparable to those distinguishing the *S. grimmia* (cytochrome *b* =2.8%; 12S rRNA=1.3%) and *C. monticola* (cytochrome *b* =6.2%; 12S rRNA=1.8%) subspecies and clearly point to an extremely close evolutionary relationship between them. The fact that *C. harveyi* and *C. natalensis* hybridizes in their regions of sympatry (Kingdon 1982) comes as no surprise. Mayr (1940) defines a subspecies as "a geographically defined aggregate of local populations which differ taxonomically from other subdivisions of the species". In the light of genetic data (present study), their phenotypic and morphological resemblance (Heyden 1968 and references therein, Ansell 1971, but see Grubb 1993, Kingdon 1982, 1997) and the hybridization evidence (Kingdon 1982), it is proposed that *C. harveyi* be relegated to a subspecies of *C. natalensis (C. n. harveyi)* as has been suggested by other authors (Ansell 1971, Heyden 1968, Ellerman *et al.* 1953, Groves & Grubb 1981).