Summer 8-2-2012

Comparative Phylogeography of Central African Duikers Using Non-invasive Sampling Methods

Stephan Ntie
University of New Orleans, sntie@uno.edu

Comparative Phylogeography of Central African Duikers Using Non-invasive Sampling Methods

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Conservation Biology

by

Stephan Ntie

BSC Université de Droit, d'Économie et Sciences d'Aix-Marseille III, Marseille, France, 2004
MS1 Université de Droit, d'Économie et Sciences d'Aix-Marseille III, Marseille, France, 2005

August, 2012

Dedication

I would like to dedicate this work to four women who impacted my life in a tremendous way. First, Ms Henriette Bisseye who was my first mother (but grandmother in a strict sense) and save literally my life when I was still an infant. Second, Ms Gisele Mifoume who is actually my real mother and did everything she could to send me to college, while she did not have the chance to even go to high school herself. Third, Dr Nicola Anthony who got me interested in Conservation Biology in 2000 at the Université des Sciences et Techniques de Masuku, Franceville, Gabon and became my PhD advisor in the same field in 2006 at the University of New Orleans, Louisiana, USA. Finally, Ms Priscilla Ntie who is my wife and has supported me since 2002 when I first met her in Marseille, France. I know that God loves me through these four women and I promise to do my best to never disapoint them.

Acknowledgement

I wish to thank all those contributors to the sample reference database and funding from several countries and institutions:

Gabon:
I wish to thank the Centre International des Recherches Médicales de Franceville (CIRMF), 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 Institut de Recherche en Ecologie Tropicale (IRET), the Centre National de la Recherche Scientifique et Technologique (CENAREST), and the Département de la Faune et de la Chasse within the Ministère de l'Économie Forestière for logistical support and/or permission to conduct this research in Gabon.
I also would like to thank the Conservators Joseph Ngowou and Allogho for permission to work in the Lopé and Monts de Cristal National Parks respectively.
I also would like to thank Dr Kathryn Jeffery for provision of research facilities and logistical support at the Station d'Etudes sur les Gorilles et Chimpanzés, Centre International de Recherches Médicales de Franceville.
I also extend my thanks to Dr Oliver Hymas and Stevens Touladjan (Wildlife Conservation Society – Libreville) for their collection of market tissue samples through the program 'Projet Gibier' in the towns of Lambaréné, Okondja and Franceville.
Special thanks to Dr Patrick Mickala from the Université des Sciences et Techniques de Masuku (USTM) and Dr Flore Koumba Pambo from CENAREST for their availability and help on numerous aspects of the present project in Gabon.

Others in Africa:
I 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.
I 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, I would also like to thank the director of the Ministerio de Agricultura y Bosques Instituto 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.
I thank Dr Deborah Pires (UCLA, CA) for her contribution in facilitating sample collection in the Republic of Congo.

United States of America:
I also wish to thank Oliver Ryder (Zoological Society of San Diego) and Mitchell Eaton (University of Colorado) for providing reference tissue samples.

Funding:
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-CIRAD).
I also earned several Graduate Student Enhancement Awards from the Department of Biological Sciences at the University of New Orleans, LA.

Table of Contents

Abstract

The present study sets out to assess patterns of evolutionary diversification in central African duikers (subfamily Cephalophinae). The sampling strategy consisted of collecting geo-referenced duiker feces across 43 sites and seven countries. However, several challenges related to the use of non-invasive samples needed to be addressed prior to large scale DNA amplification. First, the best storage method for obtaining DNA from fecal samples needed to be established. Our study revealed that while silica is best for nuclear microsatellite analyses, RNA*later* is the best storage medium for maximal mitochondrial amplification. Moreover, extracting DNA as early as possible always provided the best results. Second, since it is impossible to determine the species identity of duiker feces solely based on their morphology, a simple and reliable molecular method was needed. A tree-based approach based on ~650 base pairs of the control region amplified from reference samples was found to be the most reliable method to recover the identity of unknown samples. Third, for fine scale analyses of population genetic structure, a set of twelve nuclear microsatellites were assembled from existing bovid data. These microsatellites markers were chosen because they are very polymorphic, cross amplify among targeted taxa, co-amplify with combined markers of the same multiplex, and are powerful enough for individual identification. Patterns of mitochondrial and nuclear microsatellite variation were used to test two important hypotheses of diversification in the tropics: the Pleistocene refugia and the riverine barrier hypotheses. Analyses of historical and contemporary population genetic structure were carried out on the three most abundant species in our sampling area: the bay duiker (*C. dorsalis*), the Peter's duiker (*C. callipygus*), and the blue duiker (*P. monticola*) using mitochondrial and nuclear markers described above. These data show that (1) southwest Nigeria and southwest Cameroon comprise genetically distinct populations in *C. callipygus* and *P. monticola* species, (2) signatures of demographic expansion for all three taxa are broadly coincident with the location of hypothesized upland refugia in Gabon and Equatorial Guinea and (3) the Sanaga, Ogooué, and Sangha rivers may constitute a partial riverine barrier and/or act as fluvial refugia for duikers.

Duikers; Cephalophinae; Faeces; Species; Diagnostic; Biogeography; Central Africa

Chapter 1: Introduction

The overall goal of this thesis to develop a non-invasive strategy for genetic sampling of central African forest artiodactyls and to use these data to test two important hypotheses of tropical diversification: the Pleistocene forest refugia (Haffer, 1969) and the riverine barrier (Wallace, 1852) hypothesis. The Pleistocene forest refugia states that because of climate fluctuations during the Pleistocene, rainforest taxa were fragmented into isolated forest blocks, and that this isolation potentially leading to speciation (Haffer, 1969; Haffer, 1997). The riverine barrier hypothesis states that rivers may have constituted important barriers to gene flow that have led ultimately to divergence and formation of distinct phylogeographic clades on opposing banks (Bush *et al.*, 1992; Haberle, 2005; Kershaw *et al.*, 2007; Kirkpatrick, Fowler, 1998).

Duikers represent ideal model organisms for testing the Pleistocene forest refugia hypothesis because of their recent origin in the Plio-Pleistocene (Vrba, 1995), wide association with a range of forest habitats and high species richness relative to other non-rodent mammals (Kingdon, 1997). The fact that many of these species occur in sympatry makes it possible to use a comparative phylogeographic approach to assess the impact of changes in forest cover across several taxa. In addition, duikers are ideal study models for testing the riverine barrier hypothesis in central Africa because they favor dense shrub or forest cover (Dubost, 1983; Feer, 1989c) and have distributions that flank several major rivers such as the Ogooué and the Sanaga. Both rivers are thought to be important biogeographical barriers and have already been shown to constrain gene flow in several other central African taxa (Anthony *et al.*, 2007; Muloko-Ntoutoume *et al.*, 2000; Nicolas *et al.*, 2008; Nicolas *et al.*, 2006; Telfer *et al.*, 2003). Duikers are also ideal animal models for the present study because of the ease of non-invasive (fecal) sampling, the wide range of reliable storage methods (Idaghdour *et al.*, 2003; Murphy *et al.*, 2000; Nsubuga *et al.*, 2004; Soto-Calderon *et al.*, 2009; Wasser *et al.*, 1997) and DNA extraction protocols that have been specifically developed for fecal pellets (e.g. DNA stool minikit, QIAGEN).

Central African duikers consist of three genera: the dwarf duiker *Philantomba* (with 3 species), the monotypic savanna specialist *Sylvicapra grimmia*, and the relatively species rich genus *Cephalophus* (made up of 16 species). These forest duikers constitute a very diverse and complex group of animals that constitute an important part of the wildlife community in Africa (Kingdon, 1997; Noss, 1999; Okouyi *et al.*, 2002; Treves *et al.*, 2010). They have been traditionally hunted by local people as a valuable source of protein in their diet (Albrechtsen *et al.*, 2005; Brashares *et al.*, 2004; Carpaneto *et al.*, 2007; Fa *et al.*, 2009; van Vliet, Nasi, 2008a; Wilkie, Carpenter, 1999; Wilkie *et al.*, 2005), but they are all currently facing severe threats through illegal hunting (Barnes, 2002; Carpaneto *et al.*, 2007; Lahm, 1993c; Newing, 2001; Poulsen *et al.*, 2009; Rizkalla *et al.*, 2007; van Vliet, Nasi, 2008a; Wilkie, Carpenter, 1999; Wilkie *et al.*, 1998), habitat degradation (Laurance *et al.*, 2006b; Wilkie, Carpenter, 1999), human population increase (Muchaal, Ngandjui, 1999; Thibault, Blaney, 2003; Wilkie, Carpenter, 1999; Wilkie *et al.*, 1998), pressure from poverty (Nielsen, 2006), rise in emerging infectious and epidemic diseases ((FAO, 1999; Karesh *et al.*, 1995; Lahm *et al.*, 2007), and civilian conflicts (de Merode *et al.*, 2007; Kanyamibwa, 1998; Plumptre *et al.*, 1997).

In order to sample these species in their natural habitat, this study adopted a non-invasive sampling approach. This technique is widely practiced in studies of genetic variation in many temperate taxa (Adams, Waits, 2007a; Broquet *et al.*, 2007b; Hung *et al.*, 2004b; Prigioni *et*

al., 2006; Taberlet *et al.*, 1996; van Vliet *et al.*, 2008), but has not yet been widely utilized in the tropics. In Central Africa, challenges of collecting non-invasive samples arise because of several reasons: (1) the lack of reliable and secure infrastructures (e.g., roads) that enable access to wildlife (Minnemeyer *et al.*, 2002), (2) the specific climatic conditions of the tropics (high humidity and temperatures) that favor rapid degradation of non-invasive samples such as feces (Wasser *et al.*, 1997) without an appropriate storage method, and (3) the difficulty of identifying species (e.g., duikers) based on the appearance of their feces (Bowkett *et al.*, 2008; van Vliet *et al.*, 2008). In order to overcome these challenges, the best storage method for central African duikers needed to be found (Soto-Calderon *et al.*, 2009). In addition, an easy and reliable identification method was also necessary to discriminate between the numerous duiker species based on genetic material obtained from their feces (Ntie *et al.*, 2010a). Finally, in order to assess historical and contemporary patterns of genetic differentiation, genetic markers suited to both temporal scales were needed. For this reason, a segment of the mitochondrial control region known to be extremely variable at the population level (Ntie *et al.*, 2010a) was used to assess historical patterns of population structure and combined with a set of 12 polymorphic microsatellite loci (Ntie *et al.*, 2010b) that are more suited to contemporary population-level processes.

1.1. Hypotheses of diversification in the tropics

Among many of the hypotheses of tropical diversification that have been proposed, two of the oldest have attracted the most attention. These are: the forest refugia hypothesis of Haffer (1969) and the riverine barrier hypothesis of Wallace (1852). Both hypotheses postulate that vicariant processes leading to the isolation of related populations and subsequently to allopatric speciation (Bush, 1975; Haffer, 1969; Rosen, 1978; Wallace, 1852) The relevance of these hypotheses to tropical diversification have already been extensively assessed in South America (Brumfield, Capparella, 1996; Da Silva *et al.*, 2005; Hayes, Sewlal, 2004; Patton *et al.*, 1994) and in Australia (Joseph *et al.*, 1995; Schneider, Moritz, 1999). More recently these hypotheses have been tested on central Africa species (Anthony *et al.*, 2007; Born *et al.*, 2011; Clifford *et al.*, 2004; Muloko-Ntoutoume *et al.*, 2000; Nicolas *et al.*, 2011; Telfer *et al.*, 2003) and findings to date have shown significant effects of Pleistocene refugia and/or rivers in shaping genetic structure among several forest-associated vertebrate and plant taxa. However, their importance still remains a subject of intense debate (Bush, 1994; Colinvaux *et al.*, 2000; Haffer, 1997; Haffer, Prance, 2001; Moritz *et al.*, 2000; Schneider, Moritz, 1999).

The Pleistocene refuge theory in the tropics was first proposed by Haffer (1969) to explain patterns of species richness in the Amazon basin. The author observed that areas with high endemicity of avian fauna coincided with areas of high annual rainfall. He concluded that these areas with high annual rainfall constituted small forest blocks (forest refuges) during dry climatic periods in the Pleistocene within which forest taxa subsequently diverged from one another through drift and/or selection. The Pleistocene refuge theory has basically three predictions : (1) formation of isolated forest blocks (forest refugia) during glacial maxima that led to allopatric fragmentation and divergence between isolated populations, (2) expansion of populations during warmer interglacial periods to fill areas of suitable habitat leaving characteristic genetic signatures of population expansion, and (3) occurrence of regions of secondary contact where genetic admixture is predicted to occur between expanding neighboring refugial populations.

The most popular criticisms against the Pleistocene forest refugia hypothesis are: (a) the divergence time of many species predate the Pleistocene (Patton *et al.*, 1998); (b) it remains difficult to pinpoint the exact location of putative refugia (Moritz *et al.*, 2000); (c) its relevance to the Amazon basin is still questionable (Colinvaux *et al.*, 2000); and (d) there are numerous alternative diversification models such as the riverine barrier (Wallace, 1852) and the ecological gradient hypotheses (Endler, 1977; Endler, 1982; Schneider, Moritz, 1999) which may have played a more significant role in the process of species diversification in the tropics.

Nevertheless, many studies have been carried out to assess the existence of forest refugia, mostly in the Amazonian and Australian tropics. In contrast, relatively little attention has been granted to the African tropics, where multiple forest refugia have been hypothesized to have existed (Maley, 1996a). In addition, the Pleistocene climate was much colder and drier in west central Africa (DeBusk, 1998), favoring the formation of forest refuges, as opposed to the Amazon where it was about the same temperature but wetter (Anhuf *et al.*, 2006; Cardenas *et al.*, 2011). Examples of African refugia include the Massif du Chaillu and Mont Doudou upland in southern Gabon (Anthony *et al.*, 2007; Born *et al.*, 2011; Muloko-Ntoutoume *et al.*, 2000), Monts de Cristal in northwestern Gabon (Anthony *et al.*, 2007; Born *et al.*, 2011; Muloko-Ntoutoume *et al.*, 2000), Bamenda highlands (e.g., Mt Cameroon) (Fjeldsa, Lovett, 1997), eastern lowlands of the Congo basin (Ituri and Itombwe forests) (Fjeldsa, Lovett, 1997), and eastern African highlands and lowlands (Anthony *et al.*, 2007; Fjeldsa, Lovett, 1997; Wronski, Hausdorf, 2008). Furthermore, the existence of these Pleistocene refugia in central Africa is supported by numerous studies on phytoliths (microscopic plants remains) (Jansen *et al.*, 1989; Mercader *et al.*, 2000), macrofossils (Bonnefille *et al.*, 1995), and palynological data (Adams, Faure, 1997; Bakker, Coetzee, 1988; Bonnefille *et al.*, 1990; Bonnefille *et al.*, 1995; DeBusk, 1998; Dupont *et al.*, 2000; Elenga *et al.*, 1994; Jahns *et al.*, 1998; Jolly *et al.*, 1998; Jolly *et al.*, 1997; Kim *et al.*, 2010; Leroy, Dupont, 1997; Maley, Brenac, 1998; Marchant *et al.*, 1997; Marret *et al.*, 1999; Prentice *et al.*, 2000; Umer *et al.*, 2007; Vincens, 1991; Vincens *et al.*, 2005; Vincens *et al.*, 2007), which confirm a cooler and drier climate existed in the Pleistocene era in this region. These latter studies showed that the central African rainforest was reduced during the last glacial age of the Pleistocene (Bakker, Coetzee, 1988) and that montane forests shifted to about 1,000 meters lower in elevation, especially in East Africa (Bakker, Coetzee, 1988; Bonnefille *et al.*, 1990; DeBusk, 1998) where the climate remained cold and moist (DeBusk, 1998).

Palynological data have been used extensively in the tropics to uncover the effect of past climate on vegetation. These analyses indicate that glacial maxima led to transitions from contiguous forest to much more open forest (savanna like forest) in the Pleistocene leading to the formation of isolated forest blocks that may have isolated animals and plants for multiple generations. In contrast, the palynological studies carried out in central Africa indicate that the climate was 5-8 degrees Celsius cooler and drier throughout the Pleistocene (Bakker, Coetzee, 1988). At that time, the African rainforest is believed to have reduced in size by 84% from what we know today, as opposed to the Amazon which had its size reduced by only 54% (Anhuf *et al.*, 2006). Using both marine and terrestrial palynological data, Dupont and coworkers (2000) found a very complex scenario of vegetation succession in the west coast of Africa in an area extending from Angola to Senegal. They found that in the Pleistocene, the west coast of Africa had high grass pollen percentages, suggesting savanna and/or opening of the forest. However, rainforest did not disappear completely on these sites since they recorded montane (e.g. *Podocarpus*) and dry open forest elements, indicating both

cooler and drier conditions, respectively (Dupont *et al.*, 2000; Jahns *et al.*, 1998). This cold and dry climate of the Pleistocene era was due to the Inter-Tropical Convergence Zone (ITCZ) that brought lower sea surface temperatures from the polar region, a phenomenon known as upwelling (Servant *et al.*, 1993).

Unfortunately, palynological data do not allow us to pinpoint the exact location of forest refugia because studies of pollen usually come from areas far from these sites (lake and marine sediments). Therefore, a more reliable way to pinpoint the exact location of forest refugia would be by studying plant macrofossil and contemporary distribution of plants and animals (Bush, 1994; Fjeldsa, Lovett, 1997; Pintaud *et al.*, 2001; Schneider *et al.*, 1998). An analysis of the distribution of 178 bio-indicator species in tropical lowland rain forest in southern Cameroon (Campo Ma'an National Park) revealed that upper slopes of hills, upper altitudinal zones in the lowland forests, and riverbanks had a higher concentration of forest refuge taxa such as Begonia, Caesalpinioideae and Rubiaceae (Tchouto *et al.*, 2009). These plant taxa are considered strict and narrow range endemics because of their slow dispersal abilities and are therefore likely to indicate the presence of forest refugia where suitable conditions may have persisted during glacial maxima. Some other researchers use present day precipitation records to localize hypothetical Pleistocene refugia. The idea is that areas of high precipitation today are also the areas that were less affected by dryness of the Pleistocene epoch, which then kept enough moisture at that time to maintain forest vegetation and associated forest animals. In support of this hypothesis, satellite imagery and analysis of desert advance in the Quaternary indicate that the location of Pleistocene forest refugia corresponds to areas of present day high annual rainfall (more than 2000 mm) in west central Africa (Nichol, 1999). Similarly, a study using palaeo-climatic modelling of the Atlantic forest of Brazil based on current forest cover and palynological data supported the location of a large (Bahia) and a small (Pernambuco) Pleistocene refugia, corroborating current patterns of endemism in multiple taxa (Carnaval, Moritz, 2008).

The other important tropical diversification hypothesis is the riverine barrier model which was first introduced by Wallace (1849) to explain differences in species composition across rivers in the Amazon basin. The riverine barrier hypothesis states that rivers may constitute important barriers to gene flow that could ultimately lead to divergence and formation of distinct phylogeographic clades on opposing banks (Anthony *et al.*, 2007; Clifford *et al.*, 2004; Haffer, 1997; Jalil *et al.*, 2008; Patton *et al.*, 1994; Telfer *et al.*, 2003). However many factors could confound the effect of rivers in shaping biodiversity. Firstly, the ability (species ecology) of the studied taxa to cross a river is likely to greatly affect the potential for divergence (Ayres, Clutton-Brock, 1992; Burney, Brumfield, 2009; Lehman, 2004). Equally, the characteristics of rivers (width, depth, velocity, and seasonality), and their age and history are also important factors to consider (Ayres, Clutton-Brock, 1992; Lehman, 2004). Furthermore, in order to test this hypothesis rigorously, samples should be collected from paired sites along tested rivers from the mouth to the source (Patton *et al.*, 1998; Patton *et al.*, 1994). This is especially important because it controls for geographic distance between paired sites on opposite river banks. Historical and contemporary population structures can then be elucidated from patterns of phylogenetic and genetic differentiation at multiple points along the river's course. Consequently, since there is little to no geographical isolation in the headwater regions, where dispersal is fairly uninhibited, sister species and/or populations would harbour little to no genetic differentiation, whereas genetic differentiation would be much greater at the mouth. Therefore, cross-bank genetic differentiation should increase from headwaters to the mouth (Haffer, 1997; Patton *et al.*, 1998; Patton *et al.*, 1994). For instance, In the South American tropics, Hayes and Sewlal (2004) tested the effect of the Amazon

River as a barrier to 448 species of passerine birds from the Napo River, from its confluence to its delta. Their goal was to determine if the upper and lower sections of the river were equally effective as a barrier to the migration of these birds. They also assessed phylogenetic constraints and ecological factors believed to play a role in species' ability to cross the Amazon River. They found that the lower portion of the Amazon River was most effective as a dispersal barrier to forest specialist and *terra firme* species of birds compared to the upper reaches of the river. However, the Amazon River could not limit the dispersal of understory forest species birds, while it was a barrier to forest canopy bird species. Nevertheless, Hayes and Sewlal's (2004) findings were limited by the fact that their study only considered the present day role of the lower part of the Amazon River as a vicariant barrier but did not allow any inference about its historical role as a barrier through phylogeographic analysis. It then seems like certain taxa might be more impacted than others by river bodies, depending on their natural ability to cross them or not.

1.2. Examining the role of refugia and rivers in central Africa

In the case of African tropics, Anthony and colleagues (2007) assessed the role of Pleistocene forest refugia and rivers in shaping pattern of genetic diversity in central African gorillas (*Gorilla gorilla*). They found evidence of forest refugia in upland areas of Gabon and Equatorial Guinea and a putative lowland refugium in the Central African Republic. They also found evidence of several rivers presumably acting as barriers to gorilla gene flow. Similarly, previous studies of the rainforest tree *Acoumea klaineana* (Muloko-Ntoutoume *et al.*, 2000) and mandrill (*Mandrillus sphinx*) populations along the Ogooué River in Gabon (Telfer *et al.*, 2003) also suggested a role for Pleistocene refugia and rivers in structuring genetic diversity in Central Africa. In addition, Eriksson and colleagues (Eriksson *et al.*, 2004) evaluated the role of several rivers in the genetic differentiation of five bonobo populations in DRC. They found indeed a significant isolation by distance effect when measuring pairwise geographic distance between samples detouring rivers, as opposed to straight lines, suggesting an obvious barrier effect of these rivers on bonobo populations (Eriksson *et al.*, 2004). Although these latter studies support the ideas of forest refugia and riverine barriers in central Africa they were carried out on only one species at a time. A comparative multi-species approach is therefore needed to test for concordant patterns across multiple taxa and to evaluate whether these patterns reflect a common biogeographical history (Bermingham, Moritz, 1998; Joseph *et al.*, 1995). This is why duikers appeared as the ideal study model for testing the riverine barrier hypothesis due to the possibility of studying simultaneously several sympatric species within the whole Congo basin rainforest area (Dubost, 1984; Lahm, Tezi, 2006; Newing, 2001) and the ease with which they can be sampled non-invasively. Furthermore, all of the studies cited above have found some evidence of several rivers acting as barriers to the migration of tropical taxa but they all fail to rigorously test the riverine barrier hypothesis assumptions. In fact, these studies neither sampled in paired sites along the course of the tested rivers, nor tested if genetic differentiation increases on opposite banks from the source toward the mouth. The present study aims to address these deficiencies. Lastly, the present study aims to assess more historical patterns of genetic differentiation sequence data obtained from the hyper-variable mitochondrial control region (see Ntie et al., 2010a) whereas more contemporary patterns of population genetic structure can be assessed by using a set of 12 polymorphic microsatellite loci assembled in three multiplexes (Ntie *et al.*, 2010b).

1.3. Duiker taxonomy and ecology

Central African duikers are a species-rich group of small to medium sized mammals (~20 species) in the subfamily Cephalophinae. The name "duiker" is Afrikaans for diver and it illustrates their characteristic behaviour of diving into the underbrush when feeling threatened, using their powerful hind-legs. All forest duikers have small, back-pointing horns, which are close to the skull, allowing them to move through the thick underbrush forest without getting tangled. In general female duikers are slightly larger than males (Feer, 1989c). These antelope, which are endemic to Africa, inhabit a range of forested and woodland habitats across sub-Saharan Africa and play an important role in seed dispersersal of many plants throughout their range (Feer, 1989c). Most of them are forest specialists except the common duiker (*Sylvicapra grimmia*), which is exclusively found in savanna habitat (Kingdon, 1997).

Phylogenetic relationship within this group is still a matter of debate because of several unresolved relationships between sister taxa (Johnston *et al.*, 2011; Ntie *et al.*, 2010a; van Vuuren, Robinson, 2001). Nevertheless, three genera are generally well recognized: (1) the dwarf *Philantomba,* (2) the savanna specialist *Sylvicapra* which comprises only one species and (3) the forest dwelling *Cephalophus*. Phylogenetic analyses strongly support the monophyly of the genus *Philantomba* (van Vuuren, Robinson, 2001) and its basal position relative to other duiker taxa (Johnston and Anthony, in revision). The latter genus comprises three species (one of which was only recently described by Colyn et al., 2010) and can be distinguished from other genera by their overall morphology (small size – up to 5kg only and 36 cm tall) and karyotype (Dubost, 1980; Robinson *et al.*, 1996). The monotypic genus *Sylvicapra* falls within a well supported clade of the genus *Cephalophus,* suggesting its inclusion within it (Johnston and Anthony, in revision). Of these three genera, *Cephalophus* has the highest number of species which are further subdivided into three major lineages: (1) the giant duikers (*C. silvicultor, C. spadix, C. dorsalis,* and *C. jentinki*), (2) the East African red duikers (*C. leucogaster, C. rufilatus, C. nigrifrons, C. natalensis, C. rubidus,* and *C. harveyi*) and (3) the West African red duikers (*C. callipygus, C. weynsi, C. ogilbyi,* and *C. niger*) (van Vuuren, Robinson, 2001). The earliest fossil record of the Cephalophinae has been dated to 5.8 Ma (Vrba, 1995). A recent study (Johnston and Anthony, in revision) dated the divergence between the giant duiker and *Sylvicapra* lineage from the red duikers at 4.35 Ma, and the split between East and West African duikers at 3.08 Ma. The remaining duiker species appeared in the Pleistocene at less than 2.59 Ma (Johnston and Anthony, in revision). So, duikers appeared very recently and radiated rapidly in the Pleistocene, which explains why several sister taxa do not form monophyletic clades and cannot be discriminated from one another (Johnston *et al.*, 2011; Ntie *et al.*, 2010a; van Vuuren, Robinson, 2001). In fact, five pairs of species form paraphyletic associations: (1) *C. natalensis/C. harveyi*, (2) *C. ogilbyi/C. callipygus*, (3) *C. silvicultor/C. spadix*, (4) *C. jentinki/C. dorsalis*, and (5) *P. monticola/P. maxwelli*, (Johnston and Anthony, in revision). However, *C. callipygus* and *C. ogilbyi* are completely sympatric, while *C. nigrifrons* and *C. rufilatus* show evidence of mitochondrial introgression and extensive backcrossing. The latter case is supported by the geographic distribution of the two species which come into contact with one another at several points in their West/central African distribution (Johnston and Anthony, in revision).

Duikers are ruminant animals possessing a multiple chambered stomach that enables them to effectively digest plant tissue. They are considered to be primarily frugivorous and folivorous, with relative proportion and composition of their diet varying with season (Feer, 1989a). Indeed, they have been reported to forage on leaves, shoots, seedling, fungi and

occasionally other animals (Dubost, 1984; Emmons *et al.*, 1983; Feer, 1989a; Hofmann, Roth, 2003). In addition, duikers use their pre-orbital scent glands as a means of social communication and territory delimitation (Dubost, 1980; Dubost, 1983). In central Africa, up to seven or more duiker species are believed to occur in sympatry (Dubost, 1983; Lahm, Tezi, 2006; Newing, 2001). This may have promoted an intense competition between species and ultimately led to differentiation in body size, habitat association, diet, and timing of their activity patterns (Dubost, 1979; Feer, 1989a; Feer, 1989c; Newing, 2001; Prins *et al.*, 2006). For example, duiker body weight varies greatly (3-80 kg) and so does the size of their gape, throat, and jaw musculature. These size differences may have reduced the among species competition for the choice of food items (Dubost, 1984; Feer, 1989c; Newing, 2001). Within forest duikers, extensive studies on their ecology have been carried out on only a few species.

With respect to habitat use, the white-bellied (*C. leucogaster*) is known to make extensive use of secondary and gallery forest vegetation. The blue duiker (*P. monticola*) occupies a far broader range of habitats including dense forest, woodland mosaics and forest fragments (Dubost, 1980; Gautierhion *et al.*, 1980; Heymans, Lejoly, 1981; Kingdon, 1997). The bay (*C. dorsalis*) and Peter's duikers (*C. callipygus*) are encountered in the same forest types as *P. monticola*, though they also inhabit more dense forests (Dubost, 1980; Feer, 1989c). The black fronted duiker (*C. nigrifrons*) is exceptional since it favors flooded or swamp forest, and possesses long, flat hooves adapted to this environment (Dubost, 1984; Heymans, Lejoly, 1981). *C. silvicultor* (yellow backed duiker) is the largest of all duikers (weighting up to 80 kilograms) and has the widest distribution throughout Africa. Although its range is sparse, it has often been reported in marginal areas between the rainforest and the savanna (Kranz, Lumpkin, 1982; Lumpkin, Kranz, 1984). This species is a habitat generalist and has a diurnal activity restricted to early morning, late afternoon, and early evening (Kranz, Lumpkin, 1982; Lumpkin, Kranz, 1984). *C. rufilatus* (red-flanked duiker) is found at the edge of the forest and savanna. With respect to reproduction, *P. monticola, C. callipygus,* and *C. dorsalis* have all been found to reproduce all year long, with the peak in offspring production before each raining season when the conditions are most favorable for maximal foraging (Dubost, Feer, 1992).

The blue, Bay, and Peter's duikers are the most commonly observed in the region and have been studied extensively using radio-telemetry and mark-release-recapture by Feer (1989a,b, 1995) and Dubost (1980, 1983, 1984, 1992) in North East Gabon. These studies have provided important information on home range use, activity patterns, seed dispersal, and social organization. A detailed study of the blue duiker has shown that this species is highly territorial, diurnal, associated in breeding pairs or small family groups, and occupies home ranges of 4-6 ha (Dubost, 1980; Karesh *et al.*, 1995; Mockrin, 2010; Tutin *et al.*, 1997). In contrast, studies of Peter's (*C. callipygus*) and bay duikers (*C. dorsalis*) have shown that these species appear to be more solitary in nature, polygynous, and entirely restricted to mature forests (Karesh *et al.*, 1995; Kingdon, 1997), though the bay duiker chooses more open areas and its population is about two fold less dense than Peter's duiker (Heymans, Lejoly, 1981). In addition, the home ranges of females of both species and male *C. callipygus* are very similar in size (~40 ha), while male *C. dorsalis* have a much larger home range of 80 ha. Moreover *C. callipygus* is diurnal and *C. dorsalis* is nocturnal allowing for some temporal segregation between the two species (Feer, 1989c). As a forest interior specialist, Peter's duiker may then be the most ideal of the duiker species to test the effects of Pleistocene forest refugia on population genetic structure, whereas blue duikers are much more generalist in their use of habitat making them an interesting contrast to Peter's duiker.

1.4. Non-invasive (feces) sampling as tool in Population Genetics and Biogeography

Population genetics and biogeography studies often require collecting animal or plant samples from which DNA is extracted, amplified and molecular data are then generated for analysis. The traditional method to get these samples consists of capturing the whole specimen or collecting parts (body tissues or blood) from dead or living specimens. The latter practice usually leads to the death of the whole animal and/or difficulties in gathering information about sample identity and geographic origin, in particular when whole carcasses are not available. Fortunately, the Polymerase Chain Reaction (PCR) now allows the amplification of limited quantities of DNA (Saiki *et al.*, 1988) using short oligonucleotides (primers) specific to a given genome region of the studied taxa. In non-invasive genetics, samples are collected from hair, feces, urine, feathers, shed skin, saliva, or egg shells (Taberlet, Fumagalli, 1996; Taberlet, Luikart, 1999) left behind by the animal of interest, allowing the study of endangered, elusive, dangerous, or widespread taxa.

In the present study, duiker feces were used as source of DNA. Feces have proved to be an extremely valuable genetic resource for studies of wild animals since they are often abundant and can easily be sampled in the field without causing any harm to the animal (Kohn, Wayne, 1997b). However, only a few non-invasive genetics studies have been conducted in tropical environments, and they are mostly limited to primates (Bradley *et al.*, 2001a; Bradley *et al.*, 2002; Bradley *et al.*, 2001b; Frantzen *et al.*, 1998; Gerloff *et al.*, 1995; Nsubuga *et al.*, 2004) and carnivores (Vallet *et al.*, 2008). Non-invasive genetic studies have been used to assess population genetic structure (Broderick *et al.*, 2003; Kohn *et al.*, 1999a; Prigioni *et al.*, 2006; Valiere *et al.*, 2003), species/individual detection (Adams, Waits, 2007b; Oliveira *et al.*, 2010), sex identification (Hung *et al.*, 2004a; Kohn *et al.*, 1995; Nsubuga *et al.*, 2004; Pages *et al.*, 2009), kinship and paternity (Constable *et al.*, 2001a; Garnier *et al.*, 2001a; Hedmark, Ellegren, 2007), molecular parasitology (Boucher *et al.*, 2005a; Bretagne *et al.*, 1993a; Traub *et al.*, 2004b), and analyses of diet (Deagle, Tollit, 2007; Hoss *et al.*, 1992; Symondson, 2002b). Although non-invasive genetics is expensive and time consuming, feces were chosen as a source of DNA in the present study because: (1) duikers are very elusive animals, which makes them very difficult to catch and even get a sight of them directly in the field; (2) several target species are endangered and listed under the Convention for International Trade in Endangered Species (CITES) and protected in countries that comprise our study area; (3) this approach allows us to sample many species simultaneously over the whole Congo basin region; (4) it is easy to collect them in the field and record the precise geographic location of each individual sample; and (5) straightforward and effective methods of DNA extraction from fecal materials (e.g. DNA stool minikit, QIAGEN) are available.

Despite these many advantages of duiker non-invasive genetics, the quality and concentration of fecal DNA can be affected by environmental (e.g., humidity, temperature, UV radiation), dietary (carnivorous or herbivorous) or technical (storage type, DNA extraction protocol and size of the PCR product) factors (Broquet *et al.*, 2007a; Nsubuga *et al.*, 2004; Piggott, Taylor, 2003a; Waits, Paetkau, 2005a). In the tropics, high temperatures and humidity can accelerate the degradation of fecal DNA in terrestrial mammals (Bayes *et al.*, 2000b; Wasser *et al.*, 1997). Moreover, DNA extraction and amplification are critical steps where numerous issues can arise such as contamination from non-target DNA or errors arising from the failure to amplify a given allele (dropout), spurious amplification (false allele), or PCR failure. For instance, a study of microsatellite amplification of single shed hairs from wild chimpanzees (*Pan troglodytes verus)* revealed high allelic dropout that led to 31% of false homozygotes (Gagneux *et al.*, 1997). The author's advice was to run three separate amplifications of the

same individual or to pool hair samples from the same individual into only one. Another solution would be to adopt the multi-tube approach of Taberlet and colleagues (1996), which is the amplification of three or seven independent reactions of the same sample in order to reach a 99% probability of typing the correct genotype. A third solution involves using a quantitative PCR assay for selecting samples with sufficient nuclear DNA concentrations to recover the correct genotype (Morin *et al.*, 2001; Soto-Calderon *et al.*, 2009). This is critical because the recovery of consensus genotypes (Taberlet, Fumagalli, 1996) could be greatly compromised due to genotyping errors such as allelic dropout, false alleles, and failed amplification, as is common in the field of non-invasive genetics. Indeed, Taberlet and colleagues (1996) recommend a minimum of 56 pg of DNA to ensure 99% probability of obtaining a reliable genotype, while Morin (2001) recommended 25 pg. It is also recommended to assess genotyping errors arising from allelic dropout in order to calculate the appropriate number of independent PCR replicates needed to ensure correct genotyping (Morin *et al.*, 2001; Soto-Calderon *et al.*, 2009). Furthermore, if individuals need to be differentiated from one another, the probability of identity (PI) given the microsatellite allele frequency of typed loci should also be calculated. PI is defined as the probability of finding two individuals sharing the same genotype by chance for a given set of typed loci within a given population. PI is critical since it subsequently reveals how much missing data per sample is allowed in the dataset in order to differentiate genotypes of different individuals. Taberlet and Luikart (1999) established that at least 14 loci having each a heterozygosity of H=0.6 are required in order to reach a probability of 0.0001 of finding two siblings with identical genotypes, when dealing with non-invasive samples. However, a full-sib probability of 0.01 is generally accepted for individual identification (Viteri, Waits, 2009; Waits *et al.*, 2001).

Consequently, the use of feces as source of DNA, in our study system, raised two difficulties which needed to be overcome prior to genetic analysis. First, since there is no agreement in the literature about the best storage method for forest artiodactyl feces (Broquet *et al.*, 2007b; Frantzen *et al.*, 1998; Murphy *et al.*, 2000; Wasser *et al.*, 1997), a pilot study was needed in order to find out which of the three commonly used methods in the field is better at preserving duiker feces (Soto-Calderon *et al.*, 2009). Second, as it is impossible to discriminate species from their feces, a molecular diagnostic based on ~650 base pairs of the mitochondrial control region was developed to identify unknown fecal samples to species level (Ntie *et al.*, 2010a). Furthermore, since no one had ever genotyped duikers, 12 microsatellites loci were assembled into three multiplexes and optimized for duiker fecal DNA (Ntie *et al.*, 2010b). These data were then used to test the two main hypotheses of diversification across seven countries and 44 sites in central Africa. These data are organized into four chapters according the following specific aims:

1.5. Specific aims

Aim 1: Assess the best storage medium for short-term storage and successful PCR amplification of mitochondrial and nuclear DNA extracted from artiodactyl feces. Findings from this study will be useful for large scale sampling of artiodactyl feces in remote areas where there are no facilities for freezing for up to three months.

Aim 2: Establish a simple and efficient molecular diagnostic from a set of known reference samples in order to identify individual fecal samples to species level. A molecular diagnostic capable of recovering species identity will not only be an important first step in the process of testing the Pleistocene refugia and the riverine barrier hypotheses but also yield valuable

information on the geographic distributions and habitat associations of duiker species inhabiting Central African forests.

Aim 3: Develop a set of polymorphic microsatellite multiplexes. These nuclear microsatellites will complement the mitochondrial control region database by assessing contemporary patterns of population genetic structure across selected duiker species.

Aim 4: Use geo-referenced data to test two of the main hypotheses of tropical diversification, namely the forest refugia hypothesis (Haffer, 1969) and the riverine barrier hypothesis (Wallace, 1849). Generated data will provide critical insights into the ecological and historical processes that led to present day biodiversity patterns and in doing so might help in making the appropriate decision in conservation policy in central African rainforest ecosystems.

Chapter 2: Effects of storage type and time on DNA amplification success in tropical ungulate faeces

2.1. Introduction

Faeces have proved to be an extremely valuable genetic resource for studies of wild species because they are often abundant and can be sampled readily in the field without causing the animal undue harm (Kohn, Wayne, 1997a). Analysis of faecal DNA has been used for studies of population genetic structure (Kohn *et al.*, 1999b; Prigioni *et al.*, 2006), animal movements (Prigioni *et al.*, 2006), species/individual detection (Adams, Waits, 2007a), sex identification (Hung *et al.*, 2004b; Kohn, Wayne, 1997a; Nsubuga *et al.*, 2004), kinship and paternity (Constable *et al.*, 2001b; Garnier *et al.*, 2001b), molecular parasitology (Boucher *et al.*, 2005b; Bretagne *et al.*, 1993b; Traub *et al.*, 2004a) and diet (Deagle *et al.*, 2005; Hoss *et al.*, 1992; Symondson, 2002a).

Despite the many advantages of noninvasive genotyping, the quality and concentration of faecal DNA can be affected by environmental (humidity, temperature at time of collection), dietary (carnivorous or herbivorous) or technical (storage type, DNA extraction protocol and size of the polymerase chain reaction [PCR] product) factors (Broquet *et al.*, 2007b; Nsubuga *et al.*, 2004; Piggott, Taylor, 2003b; Waits, Paetkau, 2005b). These variables often make it difficult to disentangle the effects of such factors on DNA recovery and PCR amplification efficiency because of the lack of consistency in the sampling, storage and extraction methods used across studies (Waits, Paetkau, 2005b).

In an attempt to consider the effect of multiple storage types and storage times, Wasser *et al*. (1997) assessed faecal preservation methods in Malayan sun bears (*Helarctos malayanus*) and North American black bears (*Ursus americanus*) using up to 35 different combinations of preservation conditions and reagents. Of the many storage treatments considered, silica was found to be the best method for preserving faecal DNA for up to 6 months. These results contrast with findings from other studies, where both freezing and oven drying produced the highest amplification success rate (Murphy *et al.*, 2000). Similarly, Frantzen *et al*. (1998) found that both DMSO/EDTA buffer and 70% ethanol were the best methods for preserving faecal DNA from baboons (*Papio cynocephalus ursinus*). Finally, a recent review of noninvasive sampling methods failed to identify any effect of diet or tissue source on genotyping success (Broquet *et al.*, 2007b), bringing into question findings from previous studies.

Given the uncertainty over the best method for storing faeces before DNA analysis and the challenges of rapidly transferring samples from the field to the laboratory, pilot studies are needed in order to (i) rigorously compare commonly used storage types, and (ii) assess DNA amplification decay rates in samples stored over different time periods. Such an approach can be used to obtain samples with good amplification success and minimal allelic dropout (Morin *et al.*, 2001; Nsubuga *et al.*, 2004). In the case of chimpanzee faeces, Morin *et al*. (2001) argued that only samples with over 25 pg per reaction should be genotyped in order to ensure high genotyping success. However, it is not clear whether this approach holds for other species and under what conditions these DNA concentrations can be achieved.

In the tropics, high temperatures and humidity can accelerate the degradation of faecal DNA in terrestrial mammals (Bayes *et al.*, 2000a; Wasser *et al.*, 1997). Only a few studies comparing storage treatments or extraction methods have been conducted in tropical forest

environments, and are mostly limited to primates (Bayes *et al.*, 2000a; Frantzen *et al.*, 1998; Gerloff *et al.*, 1995; Nsubuga *et al.*, 2004; Vallet *et al.*, 2008). Since the characteristics (composition, consistency, shape and size) of faeces and type of diet are very different in primates to those of ungulates, we expect that the best storage types for maximum DNA recovery and amplification efficiency may also differ. Genetic studies using faecal samples from tropical ungulates have been carried out in a few species, including the African black rhinoceros (*Diceros bicornis*) (Garnier *et al.*, 2001a), African buffalo (*Syncerus caffer*) (Van Hooft *et al.*, 2002), African elephant (*Loxodonta africana*) (Johnson *et al.*, 2007) and warthogs (*Phacochoerus africanus*) (Muwanika *et al.*, 2007). However, none of these studies systematically assessed the effect of storage type or time on amplification success from faecal samples.

In this paper, we compared mitochondrial and microsatellite amplification success of DNA extracted from faeces of small to medium-sized ungulate species in the genus *Cephalophus* (duikers), *Tragelaphus* (spiral-horned antelope) and *Hyemoschus aquaticus* (water chevrotain). Faecal samples were stored for 1 week, 1 month and 3 months at ambient temperature in the laboratory (~23 °C) in three commonly used media: silica beads (Sigma), RNA*later* (Ambion) and 95% ethanol. We also compared nuclear DNA concentrations determined via quantitative PCR with percentage microsatellite amplification success and recovery of consensus genotypes (Taberlet *et al.*, 1996) in order to determine thresholds necessary for accurate genotyping. The effect of treatment type, storage time and nuclear DNA concentration on genotyping errors was also examined. Lastly, we also assessed whether mitochondrial amplification success could be used as a reliable predictor of microsatellite amplification success and whether there are differences in individual microsatellite marker performance over time.

2.2. Materials and Methods

2.2.1. Fecal sampling and storage methods

Species within the genera *Cephalophus, Tragelaphus* and *Hyemoschus aquaticus* are small to medium-sized ungulates that collectively make up an important component of the Central African forest ungulate community (Kingdon, 1997). *Cephalophus* spp. are the most abundant ungulate groups in Central African rainforests with up to seven species believed to occur in sympatry in the Lopé national park in central Gabon (F. Maisels, personal communication.).
These species comprise: *Cephalophus monticola* (blue duiker), *C. leucogaster* (white-bellied duiker), *C. nigrifrons* (black-fronted duiker), *C. callipygus* (Peter's duiker), *C. ogilbyi* (Ogilbyi's duiker), *C. silvicultor* (yellow-backed duiker) and *C. dorsalis* (bay duiker). Three other ungulate species are also known to occur in the same area: *Tragelaphus spekei* (sitatunga), *T. scriptus* (bushbuck) and *Hyemoschus aquaticus* (water chevrotain). Finally, it is unknown whether *Neotragus batesi* (dwarf antelope) occurs in the area since its present distribution appears limited to northern Gabon (Kingdon, 1997). As it is impossible to differentiate species by the size or shape of their faeces (Bowkett *et al.*, 2008; van Vliet *et al.*, 2008), the samples collected in the present study are likely to include representatives from multiple genera and will be referred to collectively throughout this manuscript as forest ungulates

Triplicate samples of forest ungulate faeces were taken from fresh (< 24 h) dung piles sampled over a 3-day period during the dry season in August 2005. Dung piles were sampled

along a linear transect extending through forest at the Pont de Vue (0°10′36″S11°34′39″E to 0°10′39″S11°34′15″E; *n* = 19) and Forêt de l'Aéroport (0°11′27″S11°36′8″E to 0°11′31″S11°36′31″E; *n* = 12). Three additional samples were taken in close proximity to the second transect. These sites were selected for study because they are little impacted by hunting and encompass good examples of primary and secondary lowland forest habitat favoured by many species (Emmons *et al.*, 1983). From each dung pile, two to four pellets were placed in a storage vial containing 5 g of silica gel beads (Sigma) or in a 2-mL cryovial (Fisher) containing either 1.2 mL of RNA*later* (Ambion, Inc.) or 95% ethanol. In order to avoid cross-contamination between samples, a separate leaf or a piece of wood was used to place each faecal sample into a collecting vial. Pellets were then stored at ambient temperature (~23 °C) in the laboratory for 1 week, 1 month and 3 months before DNA extraction. A pellet quantity of ~60 mg dry weight (equivalent to the pellet size of the smallest duiker *C. monticola*) was used for each extraction. DNA extractions were carried out in a room designated for noninvasive DNA work at the Centre International de Recherches Médicales de Franceville (CIRMF), Gabon, using the DNA stool minikit (QIAGEN). A blank was included in each extraction series to control for possible DNA contamination. Samples were then stored at –20 °C before PCR amplification at CIRMF or at the University of New Orleans.

2.2.2. Mitochondrial DNA amplification

Initially, the efficacy of each storage type was assessed via PCR amplification of a 300 base pairs (bp) fragment of the mitochondrial cytochrome b gene, using primers CEPHCYTBF1 (5'-TACATACACGCAAACGGAGC-3') and CEPHCYTR1 (5'-TGGAAGGCGAAGAATCGG-3'), designed to amplify across all ungulate species known to exist in the study area. PCR reactions were carried out using a GeneAmp® PCR system 9700 thermocycler (ABI) in a total volume of 20 µL containing: 2 µL of each DNA extract, 1X PCR buffer, 2.5 mM $MgCl_2$, 0.2µM of each primer, 0.5mM of each dNTP and 0.5U of *Taq* polymerase. PCR amplifications were carried out using a 3 min initial denaturation step at 94°C, followed by 35 cycles of 94°C for 30 s, 50°C for 30 s and 72 °C for 30 s, and a final extension period at 72°C for 10 min.

2.2.3. Quantitative PCR

The concentration of fecal DNA was calculated using a quantitative PCR assay based on a 96bp fragment of the tumor suppressor gene p53. Forward and reverse primers (P53iiF-GGAGCACTAAGCGAGGTGAG and P53iiR-GGAAAGAGGCAAGGAAAGGT) were designed from regions conserved across alignments of all target duiker species and a reference bovid sequence. All samples were amplified in triplicate using the SYBR Green kit (BIORAD) using the following protocol: initial step at 95°C for 3 min followed by 40 cycles of 95°C for 30 s and 62°C for 30s. DNA from a sample of *C. silvicultor* (270.9 ng/µL DNA) was selected as a standard and serial dilutions (1:50, 1:100, 1500, 1:1000, 1:5000 and 1:10000) were used to construct a standard curve, where the log of the initial DNA copy number is inversely proportional to the threshold cycle (C_T). C_T was calculated for each sample as the number of cycles where fluorescence reaches a significant threshold that is at least 10 times the mean of the standard deviation of the baseline during the first 10 to 15 cycles (Gibson *et al.*, 1996). A standard curve was generated for each run and estimation of DNA concentration was carried out using the iCycler iQ version 3.1 software (Bio-Rad). Triplicates for each treatment combination (storage type and time) for a given individual sample were simultaneously assessed in the same experiment. Only standard curves with

correlation coefficients of 98% or greater and PCR efficiency of 82% [where the efficiency = (10–1/slope – 1)*100%] or more were considered acceptable.

2.2.4. Microsatellite analysis

Existing bovid microsatellite markers were selected for this study on the basis of the following criteria: (i) amplification of a specific PCR product within the genus *Cephalophus* of the expected size range; (ii) confirmation of microsatellite repeat motifs in *Cephalophus* via DNA sequencing and (iii) within species allelic variation as verified by polyacrylamide gel electrophoresis of PCR products from at least six individuals of five *Cephalophus* species. Of the 34 published bovid markers initially assessed, eight microsatellite markers were selected to form two multiplexes: (1) INRA 40 (Beja-Pereira *et al.*, 2004), BM1225, BM2113, and BRRIBO (Bishop *et al.*, 1994), (2) BM143, BM1862, BM864 (Bishop *et al.*, 1994) and INRA05 (Beja-Pereira *et al.*, 2004). These two microsatellite multiplexes were also tested on *T. spekei* and *H. aquaticus* tissue samples in order to ensure that they amplified effectively across all potential target species. PCRs were carried out using the multiplex PCR kit (QIAGEN) in reaction volumes of 20 μL using 1× master mix, 0.4 μm of each primer and 2 μL faecal DNA. PCR conditions were as follows: an initial step at 95 °C for 15 min, followed by 35 cycles of denaturation at 94 °C for 30 s, 58 °C for 90 s and 72 °C for 60 s, and ending with a final extension at 60 °C for 45 min to ensure the addition of the 'A artefact' to the 3′ end of the double-stranded DNA. Markers INRA05 and BM864 subsequently failed to reliably amplify in the second multiplex and so were disregarded from this study, leaving a total of 6 markers for assessment of microsatellite amplification success (Table 1).

Table 1: Individual characteristics of six multiplexed microsatellites.

Locus	Size range (bp)	T_M*	Fluorescent dye	Multiplex
INRA40	132 - 199	54	Hex	1
BM1225	212 – 254	54	Ned	1
BM2113	118 - 146	54	Ned	1
BRIBBO	232 - 256	58-60	Fam	1
BM143	76 – 109	56-58	Fam	2
BM1862	193 - 217	48-60	Hex	2

*T_M represents the optimized annealing temperature, in Celsius degrees for each locus at 2.0mM $MgCl_2$.

Genotyping was carried out on a 3100 Automatic Sequence Analyser using 0.8 μL of 500 bp HD500 Rox size standard (ABI), 11.2 μL of formamide and 1 μL of PCR product per multiplex. Allele fragment sizing was performed using the GeneMapper Software version 4.0 (ABI) and binning of raw data was carried out using the program Flexibin (Amos *et al.*, 2007). Consensus genotypes were scored following Taberlet *et al.*'s (1996) multi-tube criteria: (i) A genotype was scored as heterozygous if each allele was observed in at least two independent PCR assays; (ii) an individual was declared homozygous if the same genotype was always observed in seven separate positive PCR replicates. In all other cases, genotypes were considered ambiguous.

2.2.5. Genotyping errors

Genotyping errors due to either allelic drop out (ADO) or false alleles (FA) were quantified using an approach similar to that implemented in the program GIMLET (Valiere, 2002). For those samples for which a consensus genotype for a given locus could be obtained, ADO and FA was quantified by comparing the genotype of each positive PCR against its respective consensus genotype. For a heterozygote consensus genotype, a score of one was given to a genotype that was missing one allele or another. A score of one was also given to each FA, whether in a heterozygous or homozygous condition. Heterozygote genotypes exhibiting both ADO and FA were given a score of two. Lastly, a score of two was given to heterozygous genotypes made up of two different FA.

2.2.6. Data analysis

Mean mitochondrial amplification success was calculated as the number of faecal samples that successfully amplified a 300-bp product divided by the total number of samples tested for each storage type and time period. Similarly, mean microsatellite amplification success was calculated as the proportion of all microsatellite loci that amplified for a given storage type and time period. The percentage of all microsatellite amplifications that contributed to a consensus genotype and the percentage of those that were consistent with that consensus were also compared across storage types. The average percentage ADO, FA and PCR failure across loci was computed by storage type and extraction time. Mean ADO and FA was also compared across nuclear DNA concentration categories: (0–24.9 pg/μL, 25–49.9 pg/μL, 50–99.9 pg/μL, 100–149.9 pg/μL and 150 pg/μL or greater). Finally, the minimum number of replicates required to obtain reliable genotypes with 99% confidence ($P < 0.01$) was also calculated from the mean ADO for each DNA concentration category using the formula in Morin *et al*. (2001). Genotyping error rates and consensus genotype recovery were only calculated for samples in which a consensus genotype could be obtained.

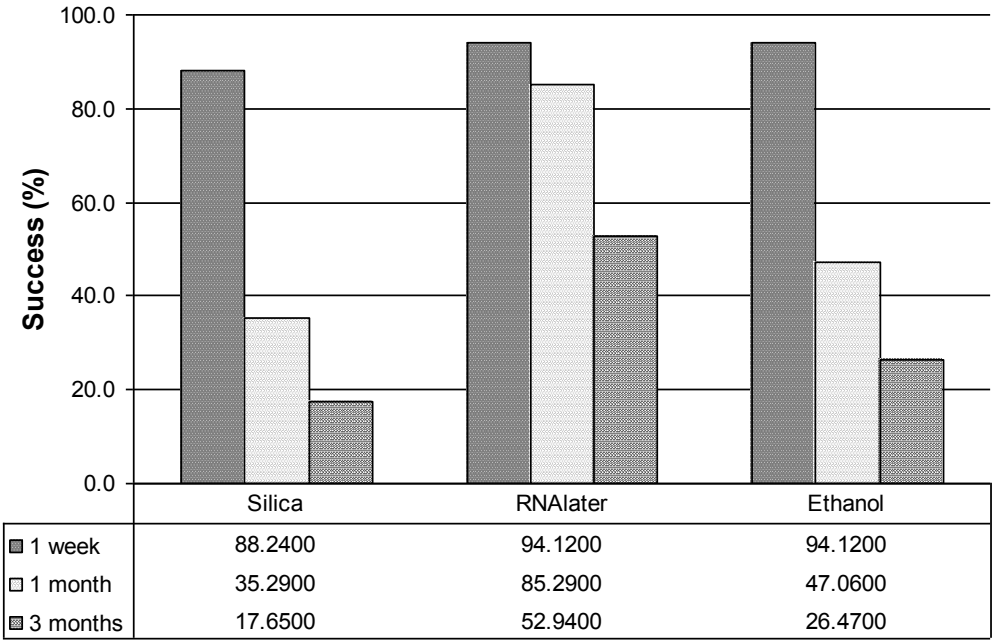
The Cochran test was used to assess the effect of storage type and time on mitochondrial amplification success. A Cochran test was also used to examine whether individual microsatellite marker performance for a given storage type differs over time. In both cases, *post-hoc* comparisons between medians were carried out using McNemar's test. As nuclear DNA concentration data and percentage microsatellite amplification success did not conform to a normal distribution, a nonparametric Friedman test was also carried out to test for the effect of storage type and time on both nuclear DNA measures and percentage recovery of consensus genotypes. In all Friedman tests, pairwise post-hoc comparisons were carried out using Wilcoxon test. A Mann-Whitney *U*-test was also used to test for differences in the mean microsatellite amplification success of samples that either succeeded or failed to amplify mitochondrial DNA. Significance of all *post-hoc* comparisons was set at the critical alpha value of 0.05 using a Holm-Bonferroni correction to control for Type I error. All statistical procedures were performed in SPSS v14.0 (SPSS Inc.).

2.3. Results

2.3.1. Mitochondrial PCR Success

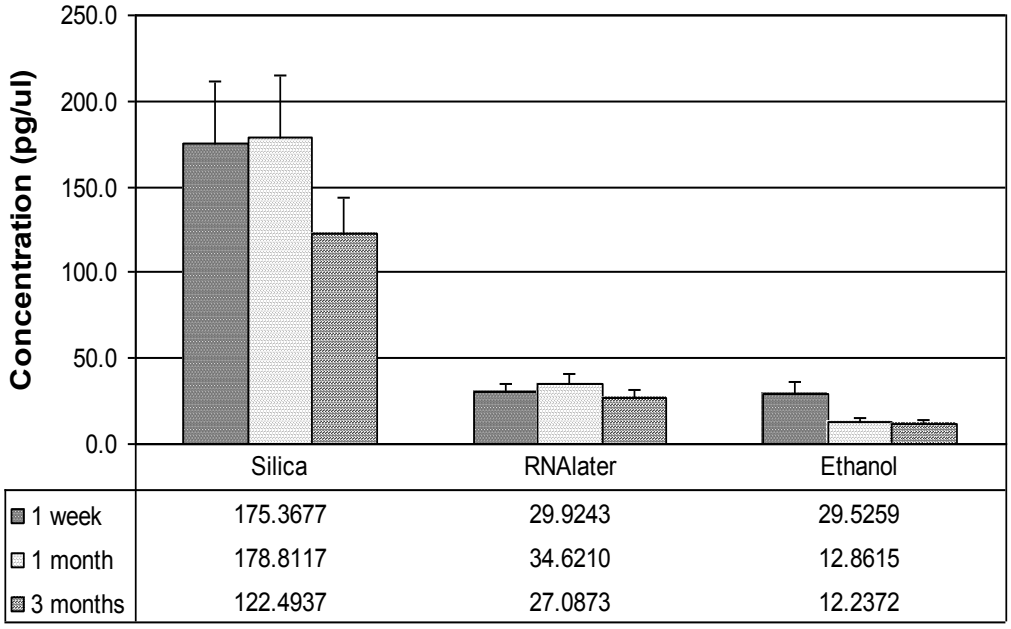
Mitochondrial amplification success differed significantly between storage types and generally decreased over time (Cochran's $Q_8 = 122.573$, $P < 0.001$; Fig. 1). When averaged over all time periods, samples stored in RNA*later* showed the highest mean success followed

by ethanol and silica. Whereas amplification success was similar across all three methods at 1 week, post-hoc tests revealed that RNA*later* outperformed silica at 1 month ($P < 0.001$) and 3 months ($P < 0.001$) of storage (Table S1, Supporting information). Amplification success in RNA*later* was also greater than ethanol at 1 month ($P < 0.001$) and 3 months ($P = 0.012$) storage. Post-hoc tests also revealed significant differences in the decay in amplification success within storage types. Significant reductions in amplification efficiency relative to the first time period were observed for both silica and ethanol at 1 month ($P < 0.001$) and 3 months ($P < 0.001$) storage. In contrast, a reduction in amplification success for samples stored in RNA*later* was only observed at 3 months ($P < 0.001$).



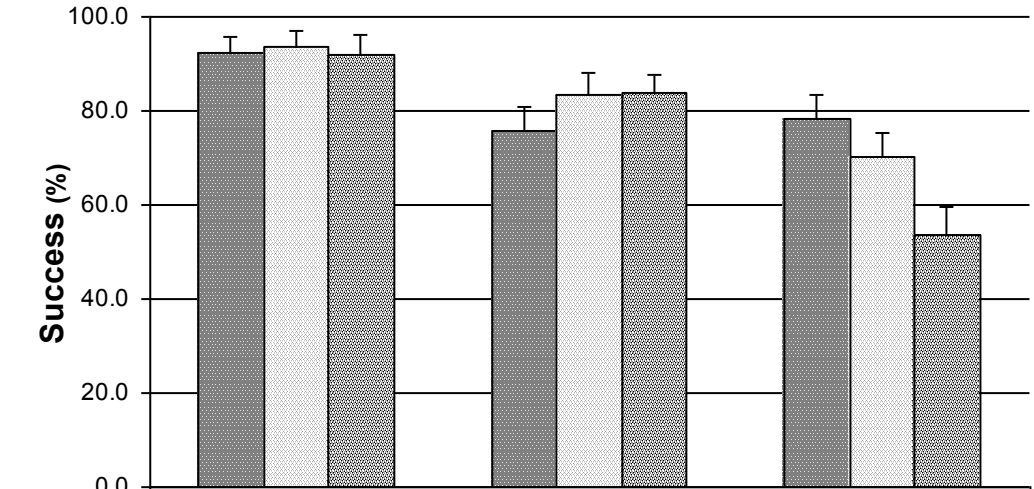
	Silica	RNAlater	Ethanol
1 week	88.2400	94.1200	94.1200
1 month	35.2900	85.2900	47.0600
3 months	17.6500	52.9400	26.4700

Figure 1. Amplification success rates for a 300bp fragment of mitochondrial DNA, from *Cephalophus* dung stored in silica, RNA*later* and 95% ethanol at ambient temperature for one week, one month and 3 months prior to extraction. Data below each figure illustrate the mean percent amplification success for each treatment.



	Silica	RNAlater	Ethanol
1 week	175.3677	29.9243	29.5259
1 month	178.8117	34.6210	12.8615
3 months	122.4937	27.0873	12.2372

Figure 2. Estimates of nuclear DNA concentration (pg/μL) in fecal samples of *Cephalophus* as determined from quantitative PCR assay. The mean concentration for each treatment is illustrated in the histogram and table below. Bars denote the standard error of the mean.



	Silica	RNAlater	Ethanol
1 week	92.5300	75.8600	78.1600
1 month	93.6800	83.3300	70.1100
3 months	91.9500	83.9100	53.4500

Figure 3. Mean amplification success rate of 6 autosomal microsatellites from fecal samples stored in 3 different media during 3 time periods. The mean amplification success for each treatment is illustrated in the histogram and table below. Bars denote the standard error of the mean

2.3.2. Nuclear DNA concentration

Significant differences in nuclear DNA concentrations between the three storage types were detected ($\chi^2_8 = 123.084$, $P < 0.001$; Fig. 2). When averaged across all time periods, silica yielded the highest overall DNA concentrations followed by RNA*later* and ethanol. Post-hoc tests revealed that DNA concentrations recovered from faeces stored in silica were always significantly higher than those extracted from samples stored in either RNA*later* ($P < 0.001$) or ethanol ($P < 0.001$) and this effect was consistent across all three time periods (Table S2, Supporting information). Excluding the first time period, samples stored in RNA*later* had higher DNA concentrations than ethanol at both 1 month ($P < 0.001$) and 3-month ($P = 0.002$) storage times. Finally, only ethanol displayed a time-dependent reduction in nuclear DNA concentration over time with a significant reduction observed at 1 ($P = 0.002$) and 3 ($P = 0.001$) months storage time.

2.3.3. Microsatellite amplification success

Storage type had a significant effect on microsatellite amplification success ($\chi^2_8 = 74.883$, $P < 0.001$; Fig. 3). Overall, samples stored in silica showed the highest mean percent amplification success followed by RNA*later* and ethanol (Fig. 3). Microsatellite amplification success in silica-stored samples was significantly greater than RNA*later* at 1 week ($P = 0.002$), 1 month ($P = 0.004$) and 3 months ($P = 0.031$) storage. Similarly, silica outperformed ethanol at 1 week ($P = 0.003$), 1 month ($P< 0.001$) and 3 months ($P< 0.001$) (Table S3, Supporting Information). Microsatellite amplification success for samples stored in RNA*later* was also significantly higher than those in ethanol at 1 month ($P = 0.012$) and 3 months ($P < 0.001$) of storage. Time dependent decay in amplification success was only observed for samples stored in ethanol. In this case, significant reductions in amplification success was observed at 3 months relative to that observed at 1 week ($P = 0.004$) and 1 month ($P = 0.018$) time periods.

Table 2. Microsatellite amplification success and genotyping errors in relation to nuclear DNA concentration (pg/μL).

DNA conc. (pg/μl)	Mean success (N ± S.E.M.)	Genotyping errors*			
		No.	ADO†	FA‡	Reps§
Silica					
0-24.9	58.33 (10, 13.21)	28	42.86	0.00	6
25-49.9	76.52 (12, 2.51)	56	25.00	0.00	4
50-99.9	95.37 (18, 1.81)	76	18.42	0.00	3
100-149.9	100.00 (12, 0.00)	61	8.20	3.28	2
150+	99.00 (29, 1.00)	163	4.29	0.00	2
RNA*later*					
0-24.9	69.86 (47, 3.91)	150	41.33	0.67	6
25-49.9	97.22 (24, 1.30)	133	21.80	3.76	4
50-99.9	92.86 (7, 3.37)	37	13.51	0.00	3
100-149.9	100.00 (3, 0.00)	16	12.50	0.00	3
over 150	-	-	-	-	-
Ethanol					
0-24.9	60.06 (58, 4.26)	167	53.89	3.59	6
25-49.9	88.24 (17, 4.23)	90	27.78	0.00	4
50-99.9	83.33 (4, 6.80)	22	13.64	0.00	3
100-149.9	-	-	-	-	-
150+	83.33 (1, 0.00)	5	0.00	0.00	-

*Genotyping errors were calculated by comparing individual genotypes to their respective consensus genotype; †allelic dropout; ‡false alleles; §number of replicate genotypes needed to reject a false homozygote ($P < 0.01$).

2.3.4. Genotyping errors

Mean microsatellite amplification success across samples increased with DNA concentration regardless of storage type (Table 2). At DNA concentrations of 50–99.99 pg/μL, ADO dropped considerably requiring only three to four PCR replicates for accurate genotyping. In contrast, the incidence of FA was always low and appeared independent of DNA concentration. ADO and PCR amplification failure was also lowest for silica samples and increased with longer storage times in all three storage types (Fig. 4). Once again, the incidence of FA appeared unaffected by storage type or time.

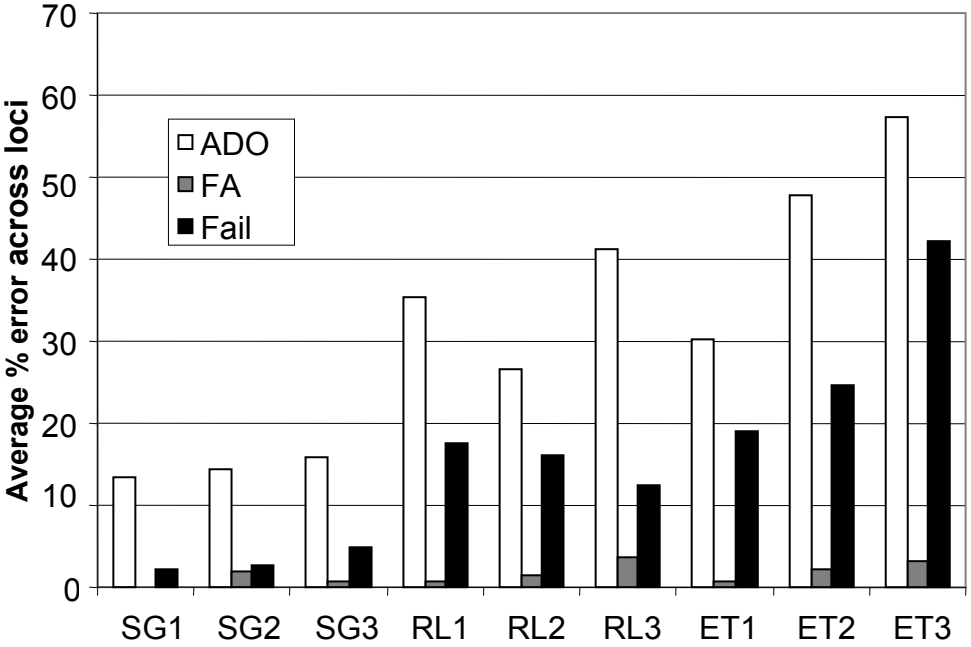


Figure 4. Percentage error rates averaged across loci for samples stored in silica gel (SG), RNA*later* (RL) and ethanol (ET) for 1 week (1), 1 month (2) and 3 months (3). Error rates are categorized by allelic dropout (ADO), false alleles (FA) and amplification failure (FAIL).

Of those samples for which a consensus genotype could be obtained, a greater proportion of positive PCR amplifications were obtained from samples stored in silica (38.04%) as compared to RNA*later* (33.70%) or ethanol (28.26%). Similarly, the percentage of these positive amplifications that recovered the consensus genotype was greatest for silica (85.24%), relative to either RNA*later* (66.67%) or ethanol (54.49%). This effect of storage type on the recovery of the consensus genotype was significant (= 111.548, $P < 0.001$) with significantly higher proportion of consensus genotypes obtained for silica than for either RNA*later* ($P \leq 0.005$) or ethanol ($P < 0.001$) across all time periods (Table S4, Supporting information). Of the latter two storage types, RNAlater yielded significantly better recovery of the consensus genotype than ethanol at either 1 month ($P < 0.001$) or 3 months ($P < 0.001$) of storage before extraction. Only ethanol showed a time-dependent decay in the number of genotypes consistent with the consensus at 1 month ($P = 0.004$) and 3 months ($P < 0.001$).

2.3.5. Microsatellite vs. mitochondrial DNA amplification success

Although the overall mean microsatellite amplification success was higher in samples that amplified mitochondrial DNA compared to those that did not, this relationship was only found to be statistically significant for samples stored in ethanol ($Z = –2.538$, $P = 0.011$).

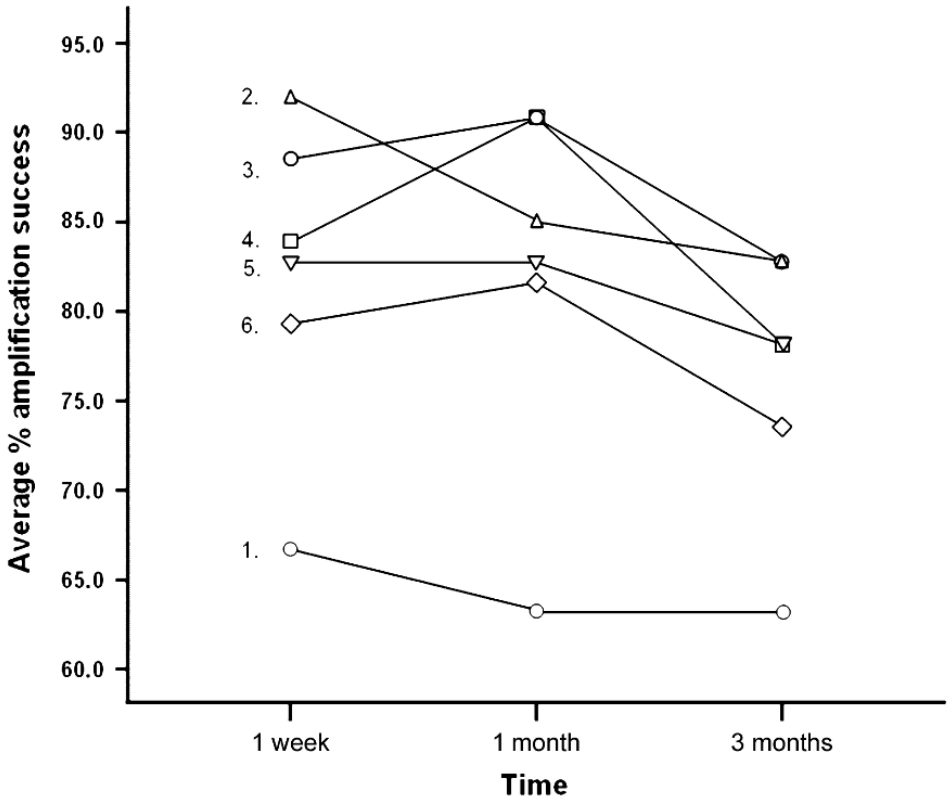


Figure 5. Average microsatellite amplification success for samples stored for 1 week (1), 1 month (2) and 3 months (3) before extraction. Amplification success was assessed using the following six loci, by increasing size: BM143 (1), BM2113 (2), Inra40 (3) BM1862 (4), BM1225 (5), BRRIBO (6).

2.3.6. Individual marker performance

With the exception of locus BM143, microsatellite amplification success decreased with increasing amplicon size (Fig. 5). Although the effect of marker performance over time was significant for samples stored in silica (Cochran's $Q_{17} = 33.378$, $P = 0.010$), RNA*later* (Cochran's $Q_{17} = 57.873$, $P < 0.001$) and ethanol (Cochran's $Q_{17} = 67.552$, $P < 0.001$), this effect is largely due to differences in amplification success between markers and not differences in decay rates of individual markers (Table S5, Supporting information).

2.4. Discussion

In general, mitochondrial and nuclear DNA extracted from fresh Central African forest ungulate faeces amplifies well in samples extracted within the first week after collection. After this period, mitochondrial amplification efficiency declines quite rapidly in storage media other than RNA*later*. In contrast, nuclear DNA appears relatively stable in both silica and RNA*later* although DNA concentrations were always much higher in the former. The greater DNA concentrations in silica-stored samples serve to maximize the percentage recovery of consensus genotypes and minimize genotyping error rates and amplification failure. Comparison of percentage ADO by DNA concentration category and storage type indicate that even concentrations as low as 25–49.99 pg/μL require only four replicates to obtain an accurate genotype with 99% certainty. In order to directly compare our quantitative PCR results with those of Morin *et al*. (2001), it is first necessary to convert DNA quantities per reaction to pg/μL concentrations. Morin *et al*. (2001) used 2 μL DNA per reaction so that

template quantities of 101–200 pg/reaction can be equated to a template concentration of 50.5–100 pg/μL. In the present study, this concentration yields an ADO rate of only 13.51–18.42% compared to 26% in the study by Morin *et al*. (2001). Accordingly, only three replicates are required for accurate genotyping at this concentration, whereas four are required in the companion study by Morin *et al*. (2001). More importantly, most samples preserved in silica in the present study have DNA concentrations of 50 pg/μL or more suggesting that many of the samples stored this way will yield accurate genotypes with few replicates. This finding is also reflected in the study by Morin *et al*. (2001) where despite longer storage times of 2–18 months on silica, the majority of samples had DNA concentrations of 50 pg/μL or higher. In contrast, most samples stored in RNA*later* and ethanol had DNA concentrations less than 50 pg/μL necessitating a greater number of replicates for accurate genotyping.

Although nuclear DNA concentrations and microsatellite amplification success generally decline with time, this effect is only significant for samples stored in ethanol. These findings then suggest that for the other two storage media the rate of nuclear decay is much less marked than that of mitochondrial DNA. One potential explanation for this slower rate of decay in the nucleus might be due to the fact that nuclear DNA is better protected against physical damage and the action of DNAses due to strong DNAhistone interactions within the nucleus (Fernandez, Vrba, 2005). Differential rates of mitochondrial vs. nuclear decay have been observed in tissue with induced physical damage, such as maceration or freezing/thawing (Berger *et al.*, 2001; Foran, 2006) and may partly explain the slower rates of nuclear DNA degradation observed here (Berger *et al.*, 2001; Fernandez, Vrba, 2005; Foran, 2006).

Another factor that could potentially have influenced this discrepancy in decay of mitochondrial vs. nuclear DNA is the inverse relationship frequently observed between the size of the targeted DNA and its amplification success rate (Broquet *et al.*, 2007b). As the size of the mitochondrial (300 bp) product in the present study was larger than that targeted from the nucleus (76–255 bp), it is possible that a systematic upward bias in nuclear amplification success was introduced, that is, the slower rate of nuclear decay is due to the shorter fragment length of our nuclear markers. Although microsatellite amplification success is generally inversely related to amplicon size, longer fragments did not appear to decline more rapidly over time. This finding suggests that amplicon size *per se* was not driving differences in rates of mitochondrial vs. nuclear decay. Differences in amplification success across taxa could also lead to discrepancies in mitochondrial vs. nuclear amplification success. This is unlikely since both the mitochondrial and microsatellite multiplex assemblies cross-amplified *T. spekei* and *H. aquaticus* reference samples, suggesting that there should be no taxonomic bias. However, the five mismatches in the p53 reverse primer-binding region of *H. aquaticus* could potentially lead to an under estimation of DNA concentrations for this species and may necessitate the design of a species-specific assay. Fortunately, the mitochondrial control region of *H. aquaticus* has a major deletion relative to *Cephalophus* and *Tragelaphus* sequences (Ntie *et al.*, 2010a) making it possible to readily identify and remove these samples. Moreover, the observed primer mismatches are unlikely to introduce a systematic bias in our data since we do not expect primer binding efficiency to be affected by either storage treatment or time.

Slower nuclear DNA decay rates may also increase the risk of amplifying nuclear translocated mtDNA (Numts) over the true mitochondrial copy. The same effect has been observed in elephants, where primers designed to amplify mitochondrial DNA preferentially

amplified nuclear copies in DNA extracted from hair whereas only mitochondrial sequences were retrieved from blood (Greenwood, Paabo, 1999). Although, we did not directly address this question in the present study, it would be interesting in future studies to examine the percentage amplification success of Numts in faecal samples stored in different treatment combinations.

In conclusion, faecal samples appear to be a good source of nuclear DNA for microsatellite genotyping, even after long periods of storage at room temperature. With the exception of ethanol stored samples, mitochondrial DNA amplification success correlates poorly with nuclear amplification efficiency and should not be used as predictor of nuclear genotyping success. In contrast, quantitative PCR is an accurate predictor of microsatellite genotyping success, and despite the high cost, remains the only reliable method of predicting high rates of nuclear genotyping success and accuracy.

The results of the present study also show that the best storage type may depend on which DNA type (mitochondrial or nuclear) is needed for downstream analyses. Silica gel is recommended for nuclear microsatellite analyses whereas RNA*later* may be better for mitochondrial DNA studies. Ethanol was the least suitable storage method and should not be used for the long-term preservation of tropical ungulate faeces. As there may be no simple answer to which method is better in other species, pilot studies should be conducted before any large-scale analyses are carried out.

Several questions arising from this study still remain unanswered. Field researchers frequently store faeces in a freezer for various periods of time before DNA extraction. This strategy has been shown to improve sample preservation and this is especially true in tropical areas where high temperatures and humidity may accelerate decomposition rates (Bayes *et al.*, 2000a; Wasser *et al.*, 1997). Further work should examine how storage time in the freezer affects patterns of amplification success observed in the present study. Systematic studies on how diet across different trophic groups (carnivores, omnivores) affects amplification success in tropical species may also ultimately help establish whether any general patterns are evident and whether these could be used to develop generic guidelines for the long-term storage and efficient amplification of DNA.

Chapter 3: A molecular diagnostic for identifying central African forest ungulates from fecal pellets

3.1. Introduction

Forest artiodactyls are an important component of Central African rainforest vertebrate communities and are widely distributed across the Congo basin (Feer, 1989a; Feer, 1989c; Kingdon, 1997; Tutin *et al.*, 1997). Many species are also an important source of protein and income for people living in the Congo basin (Fa *et al.*, 2002; Lahm, 1993c; Wilkie, Carpenter, 1999). For example, Wilkie and 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 and Blaney (2003) found in surveys conducted between 1997-8 in Gamba market of Gabon that artiodactyls make 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, 1993a). As a result, many African artiodactyls have become increasingly threatened by hunting pressure (Laurance *et al.*, 2006b; Newing, 2001; Wilkie, Carpenter, 1999). Wildlife declines are also likely to be exacerbated by the rise in emerging infectious and epidemic diseases (FAO, 1999; Karesh *et al.*, 1995; Lahm *et al.*, 2007). Demographic projections suggest that if current levels of over-exploitation persist, several important forest artiodactyls will disappear within the next one hundred years (Barnes *et al.*, 2002; Lahm, 1993b).

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 (Fong *et al.*, 2007; Moore *et al.*, 2003; Wasser *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, richness and abundance estimates of tropical forest artiodactyl species are traditionally based on line transect sampling of faecal piles (Ellis, Bernard, 2005; Walsh, White, 1999; White, 2000), direct observations (Heydon, Bulloh, 1997; Lahm, 1993a; White, 2000) and market sampling (Crookes *et al.*, 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 (Croes *et al.*, 2007; Feer, 1989c; Focardi *et al.*, 2002; Lahm, 1993a; Walker, 2006). Lastly, practical identification of bushmeat samples may also be difficult when whole animal carcasses are not available. Furthermore, 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, 1993b; van Vliet, Nasi, 2008a) and data on species geographic origin may be difficult to obtain (cf. Moore *et al.*, 2003). Since numerous similarly-sized artiodactyl species can occur in sympatry in central African tropical forests (Dubost, 1980; Dubost, 1984; Kingdon, 1997; Lahm, Tezi, 2006; Newing, 2001), 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 populations or unidentified tissues, even when only degraded sources of DNA are available (Baker, Palumbi, 1994; Dreher *et al.*, 2007; Kohn *et al.*, 1999b; Taberlet *et al.*, 1996). 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 (Baker *et al.*, 2006; Baker, Palumbi, 1994; Hammond *et al.*, 2001; Lorenz *et al.*, 2005; Ross *et al.*, 2003). This only becomes a viable strategy once (a) a substantive database of representative sequences from each target species has been obtained (Ekrem *et al.*, 2007) and (b) 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: (i) reciprocal monophyly (Wiens, Penkrot, 2002) or less commonly (ii) inter-specific genetic distances that are 10-fold or greater than the average intra-specific 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 (DeYoung, Honeycutt, 2005; Kohn *et al.*, 1999b), 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 untested.

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 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 previously diagnosed 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.

3.2. Methods

3.2.1. 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, C. ogilbyi*) and other species in the sympatric genera *Neotragus* (*N. moschatus*), *Tragelaphus* (*T. spekei, T. scriptus*) and *Hyemoschus* (*H. aquaticus*). Since 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, S. grimmia, C. adersi, C. maxwelli, C. niger, C. zebra, C. weynsi, C. rubidus and C. jentinki*. Reference tissue samples were donated from zoos, museums, scientific collections and bush meat market surveys conducted in collaboration with the Wildlife Conservation Society (WCS) 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, tissue or teeth samples obtained from M. Colyn (n=25) or from the San Diego Zoo (n=18) were accompanied by either craniometric data (Colyn, unpublished) 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: (i) Ossele (Oss), Ipassa (IP), Lopé National Park (Lope) and Monts de Cristal National Park (MCR) in Gabon; (ii) Nouabalé-Ndoki National Park (ND) in the Republic of the Congo; (iii) Salonga National Park (SA) in DRC; (iv) Bioko island (BKO) in Equatorial Guinea; (v) Udzungwa Mountains (AB), in Tanzania. Details on the geographic location and donor(s) of these samples are listed in Supplementary Table 1.

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 (BIO 101). 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 more cusps with a sterilized drill and then incubated overnight at 65°C in lysis buffer provided in the blood extraction minikit (Qiagen). DNA extraction was then carried out according to 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 DNA stool minikit (Qiagen). A blank was included in each extraction series to control for DNA contamination.

Table 1. Samples sequenced in the present study.

	No. of Individuals			
Species	CR	Cyt b	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 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 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*. (2008a)
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.

3.2.2. 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 (Paabo, 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 *et al.*, 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 50µl reaction volume containing 1X enzyme buffer (200mM Tris pH 8.4, 500mM 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) and 1-2 µl of template DNA. DNA amplifications were carried out using the GeneAmp® 9700 thermocycler (ABI) or I-cycler (BIORAD) with an initial denaturation time of 94˚C for 4 min, followed by 35 cycles of initial denaturation at 94˚C for 30s, annealing at 50-55˚C for 30s, and an extension step at 72˚C for 30s. The program 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 (BIO 101) according to manufacturer's instructions. Sequencing was carried out on an automated DNA sequencer (ABI 3100) using the BigDye Terminator Cycle Sequencing Kit V1.1 (ABI).

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

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

3.2.3. Phylogenetic analysis

Sequence data was edited using the program SEQUENCHER v 4.1.1 (Gene Codes Corporation) and aligned prior to 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 largely 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 control region alignment containing a large insert in some *C. monticola* sequences was also deleted prior to phylogenetic analysis to determine whether this affected the overall tree topology 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 v.3 (Kumar *et al.*, 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 employed 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 not defined *a priori*. In order to ensure that the MCMC chain had not been trapped in local optima (Leache, 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 greater than 500 for all model parameters. Convergence across analyses was assessed by verifying whether different runs attained the same stationary distribution and average log likelihood values. Chains were run for 10,000,000-50,000,000 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 $\geq$ 0.95.

3.2.4. 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 ten-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* (van Vuuren, Robinson, 2001). Pair-wise Kimura-2-parameter genetic distances were calculated using

the program MEGA v.3 (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 ten-fold genetic distance rule.

3.2.5. 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). 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 base pairs in length and were commercially available through New England Biolabs or Invitrogen. The reference database used to screen for RFLPs was identical to that used to build phylogenetic trees for faecal sample identification.

3.3. Results

3.3.1. 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 (Figure 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* and (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 previously observed (Moodley *et al.*, pers. com). Due 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 (Supplementary Figure 1) and MP (Supplementary Figure 2) 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.

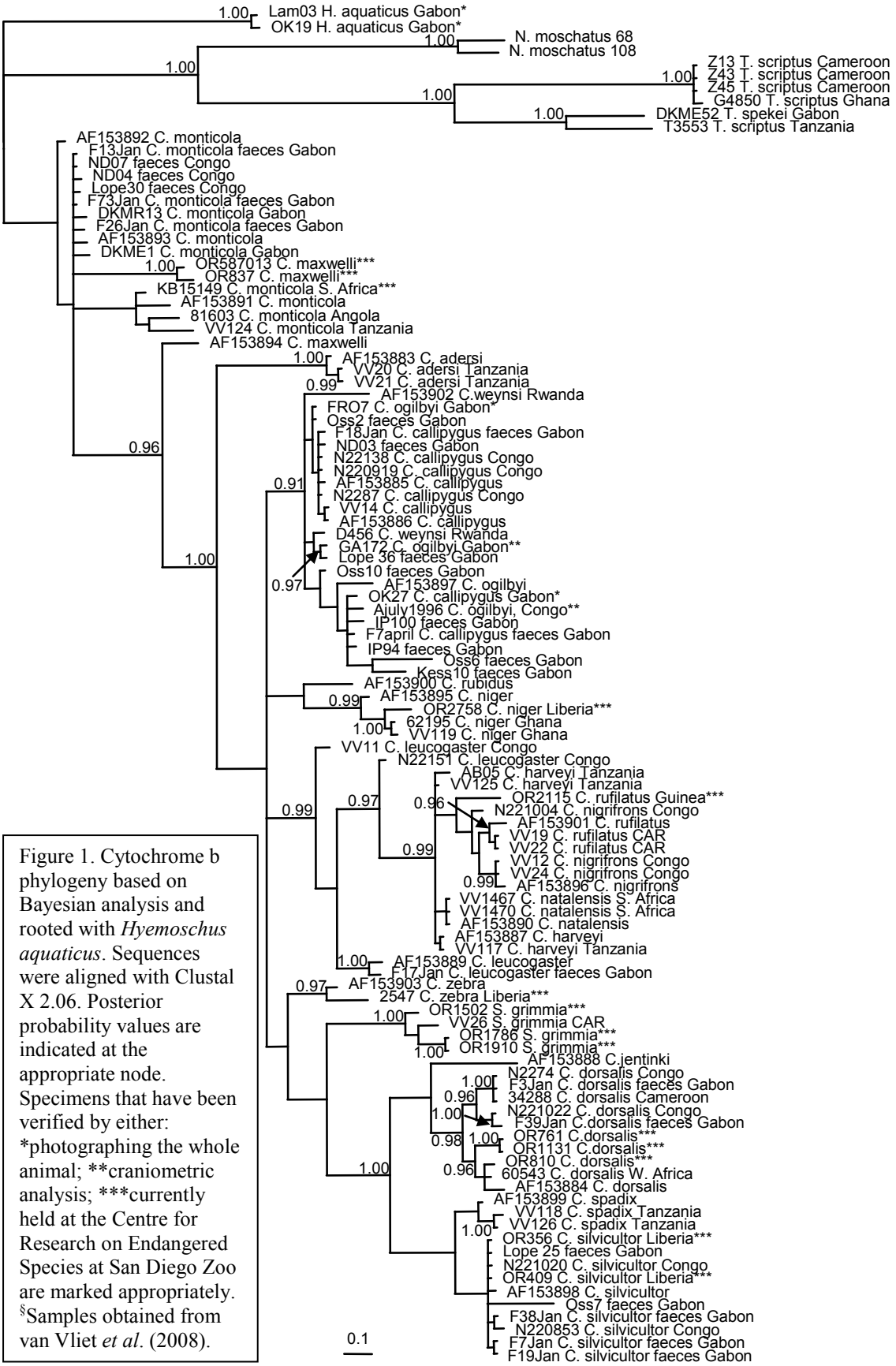


Figure 1. Cytochrome b phylogeny based on Bayesian analysis and rooted with *Hyemoschus aquaticus*. Sequences were aligned with Clustal X 2.06. Posterior probability values are indicated at the appropriate node. Specimens that have been verified by either: *photographing the whole animal; **craniometric 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).

In contrast, Bayesian analyses of the control region dataset using either alignment procedures were not only able to successfully resolve target taxa but did so with much higher posterior support (Figure 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 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* which appeared paraphyletic to *C. maxwelli* and (2) *C. harveyi* which appeared paraphyletic to *C. natalensis*. Two geographically restricted species (*C. rubidus*, *C. jentinki*) also failed to amplify, possibly due to the degraded condition of these samples. 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 (Supplementary Figure 3) or excluded (Figure 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 (Supplementary Figure 4, 5). 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.* (2008a).

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 (Figure 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. 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.

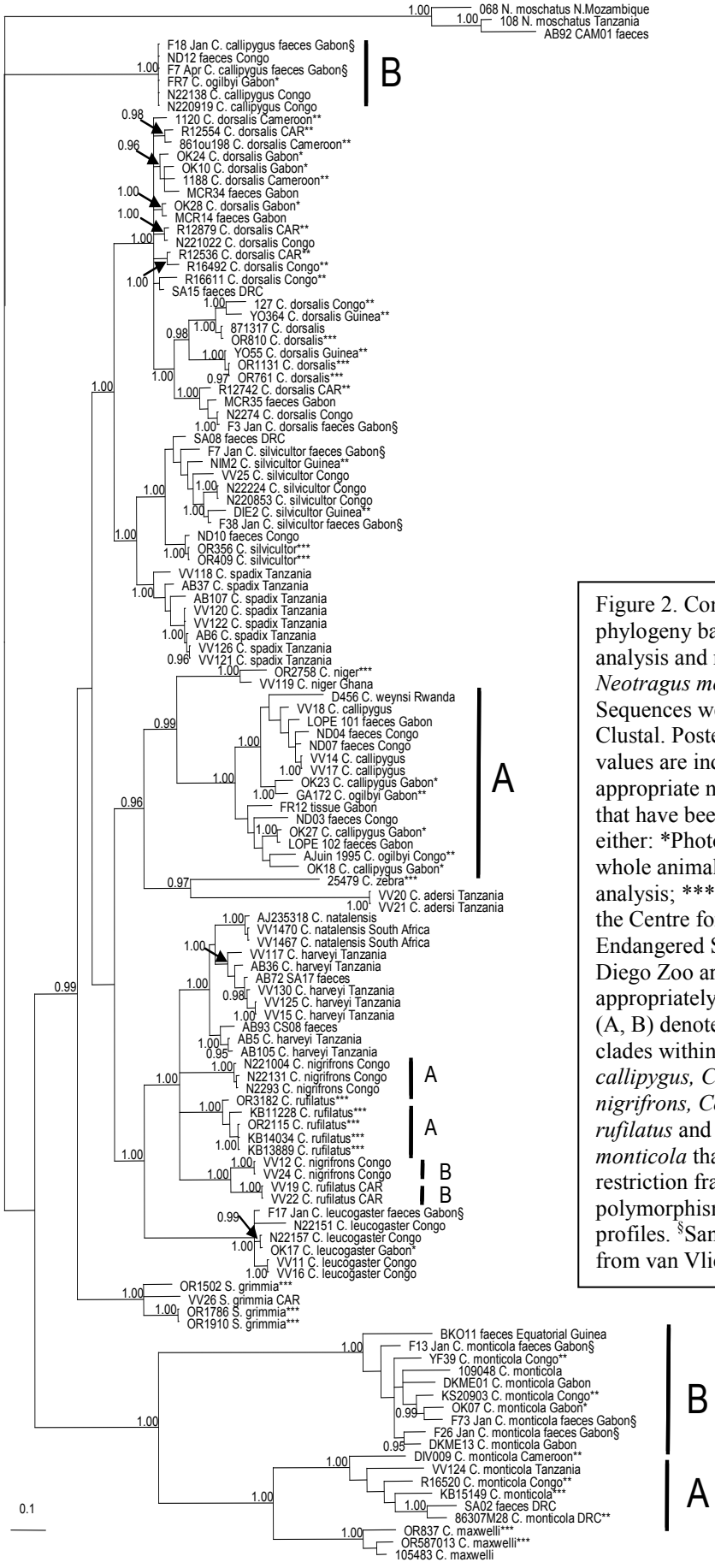


Figure 2. Control region phylogeny based on Bayesian analysis and rooted with *Neotragus moschatus*. Sequences were aligned with Clustal. Posterior probability values are indicated at the appropriate node. Specimens that have been verified by either: *Photographing the whole animal; **craniometric analysis; ***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).

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.* (2008a) was most likely mislabelled since it was originally diagnosed as *C. nigrifrons* in the 12S ribosomal RNA phylogeny of van Vliet *et al.* (2008a) but was indistinguishable from a *C. callipygus* sample (F18 Jan) included in the present analysis. Lastly, there was one sequence anomaly in the cytochrome *b* database where two different haplotypes were obtained from the same sample of *C. weynsi* (Figure 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 (Figure 2).

3.3.2. Evaluation of DNA barcoding criteria

The NJ tree constructed from Kimura-2-parameter distances for the cytochrome *b* dataset (Supplementary Figure 1) 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 ten-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).

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																			
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*	0.07	0.078	0.074	0.053	0.038	0.041	0.042	0															
9. *C. rufilatus*	0.088	0.094	0.076	0.072	0.059	0.06	0.058	0.063	0.013														
10. *C. nigrifrons*	0.08	0.087	0.067	0.065	0.053	0.054	0.053	0.055	0.015	0.008													
11. *C. natalensis*	0.077	0.09	0.08	0.062	0.053	0.053	0.053	0.051	0.023	0.022	0												
12. *C. harveyi*	0.081	0.094	0.083	0.065	0.059	0.06	0.06	0.057	0.018	0.021	0.006	0.001											
13. *C. leucogaster*	0.07	0.079	0.069	0.058	0.057	0.059	0.051	0.054	0.037	0.036	0.024	0.03	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*	0.05	0.062	0.057	0.055	0.05	0.048	0.045	0.049	0.06	0.051	0.049	0.053	0.05	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	0.08	0.048					
19. *C. monticola*	0.081	0.089	0.093	0.086	0.09	0.091	0.084	0.086	0.093	0.083	0.089	0.093	0.083	0.09	0.073	0.083	0.086	0.054	0.033				
20. *H. aquaticus*	0.144	0.16	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*	0.14	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*	0.14	0.148	0.159	0.14	0.15	0.155	0.153	0.134	0.152	0.148	0.137	0.144	0.14	0.141	0.131	0.165	0.152	0.148	0.147	0.148	0.182	0.102	0

Within species distances are on the diagonal and in BOLD.

* Denotes an estimate based on only one sample.

3.3.3. Cytochrome b and control region RFLP analysis

A cytochrome *b* gene RFLP diagnostic was developed (Supplementary Table 2) and a flow chart was designed from the combination of enzymes that together could be used in species identification (Supplementary Figure 6). However, enzyme digests yielded inconsistent results with unknown samples or tissue samples suggesting that within 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 revealed several candidate restriction enzyme sites that could be used to distinguish species (Supplementary Table 3). 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 ~275bp 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 ~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.

3.4. 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 *et al.* (2008b). This 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 closely related 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 van Vuuren and 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 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 (van Vuuren, Robinson, 2001; Vrba, 1995). 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 deeper level 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 this short fragment. 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 (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 *et al.*, 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 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 (Hickerson *et al.*, 2006; Moritz, Cicero, 2004). Pair-wise comparisons of Kimura-2-parameter distances between taxa indicate that almost all *Cephalophus* species do not satisfy the ten-fold genetic distance rule described by Hebert *et al.* (2004). This is due to both the high intra-specific diversity present in many species and their recent divergence from one another. The few cases where the ten-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 (SSCP) analysis 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 (Orita *et al.*, 1989). 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 (Vazquez-Dominguez *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 (Bowkett *et al.*, 2008; Croes *et al.*, 2007; Focardi *et al.*, 2002; van Vliet *et al.*, 2008; Waltert *et al.*, 2006), we therefore recommend using this marker to accurately recover species identity. However, while the control region phylogeny can be used to recover species identity this approach should not be used to evaluate taxonomic relationships within this group. Additional data will be used to revise the taxonomy of this group since there is 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: (i) mapping habitat preferences and species distributions, (ii) identifying bush meat samples from urban and regional markets, (iii) estimating species abundance when combined with individual multi-locus genotype data, (iv) predicting species responses to habitat modification through logging, fragmentation and intensive hunting and (v) determining species presence and relative abundance in environmental impact assessments (EIAs) of development projects and associated wildlife monitoring programs. Due 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.

Chapter 4: Cross-species amplification of bovid microsatellites in central African duikers (genus *Cephalophus*) and other sympatric artiodactyls

4.1. Introduction

Central African duikers (genus *Cephalophus*) and sympatric artiodactyl species *Tragelaphus spekei* (sitatunga) and *Hyemoschus aquaticus* (water chevrotain) constitute an important group of mammals in the Congo basin rainforest community (Feer, 1989b; Feer, 1989c; Kingdon, 1997; Tutin *et al.*, 1997). Despite their widespread distribution and importance to local people as a source of protein and income (Fa *et al.*, 2002; Lahm, 1993a; Thibault, Blaney, 2003), little is known about their relative abundance and genetic structure in the wild (Bowkett *et al.*, 2009; Ntie *et al.*, 2010a; van Vliet, Nasi, 2008b; van Vliet *et al.*, 2008).

With recent advances in non-invasive methods, multilocus microsatellite genotyping from animal dung or hair has emerged as a very efficient and reliable method for assessing wildlife population abundance and genetic differentiation (Kohn *et al.*, 1999b; Taberlet *et al.*, 1996; Taberlet, Luikart, 1999). Such multilocus applications can also be extended to species identification (Pilot *et al.*, 2007; Vazquez-Dominguez *et al.*, 2001), wildlife forensics (Wasser *et al.*, 2007), relatedness (Banks *et al.*, 2002; Lorenzini *et al.*, 2004), and to make inferences about species ecology and behaviour (Garnier *et al.*, 2001a).

Numerous bovid microsatellites have been tested and amplified across multiple taxa (Beja-Pereira *et al.*, 2004; Cosse *et al.*, 2007; Kim *et al.*, 2004; Maudet *et al.*, 2001). In a recent paper, Soto-Calderón *et al.* (2009) was able to amplify and successfully genotype unknown fecal samples from African forest artiodactyls using six published bovid microsatellite loci. However, no one has yet (a) systematically assessed the ability of a larger number of microsatellites to cross-amplify in a wide range of African forest artiodactyls or (b) assessed their potential for fine-scale studies of population genetic structure.

4.2. Methods

Ear-punch samples from harvested central African duiker species *Cephalophus monticola* (blue duiker, n=20), *C. callipygus* (Peter's duiker, n=20), *C. dorsalis* (bay duiker, n=6), *C. nigrifrons* (black-fronted duiker, n=6), *C. leucogaster* (white-bellied duiker, n=6) and *C. silvicultor* (yellow-backed duiker, n=3) were collected from logging villages (Ndoki 2, Kabo) in the Republic of Congo (Supplementary Table 1). Samples were stored in a NaCl saturated buffered solution containing 0.5 M EDTA (pH 8.0) and 20% DMSO at room temperature in the field and then at -20°C once in the lab. Ear, tail, or muscle tissue samples of *T. spekei* (sitatunga, n=6), *Hyemoschus aquaticus* (water chevrotain, n=6) and *C. silvicultor* (yellow-backed duiker, n=3) were also collected from Okondja, Lambaréné and Franceville in Gabon (Supplementary Table 1) and stored in 95% ethanol. DNA was extracted from ~100mg tissue using a standard phenol-chloroform extraction protocol (Sambrook, Russell, 2001) and re-suspended in 200µL of Tris-EDTA buffer (10 mM Tris-Cl, 1mM EDTA, both at pH 8.0). A blank was included in each extraction series to control for DNA contamination.

A total of 34 microsatellite loci with demonstrated polymorphism in other bovid species were tested for their ability to cross-amplify in six duiker species: *Cephalophus callipygus, C. monticola, C. silvicultor, C. nigrifrons, C. dorsalis* and *C. leucogaster* (Table 1). Three to six samples of each species were amplified and their polymorphism was assessed using polyacrylamide gel electrophoresis. Published primer pairs for these loci were used (Barendse

et al., 1994; Beh *et al.*, 2000; Bishop *et al.*, 1994; Buchanan, Crawford, 1993; Crawford *et al.*, 1995; Kaukinen, Varvio, 1993; Kogi *et al.*, 1995; Mommens *et al.*, 1994; Penty *et al.*, 1993; Steffen *et al.*, 1993; Vaiman *et al.*, 1994), with the exception of locus SR12 where an internal reverse primer (R3) was designed to yield a smaller amplicon. PCR amplifications were carried out with a GeneAmp ® 9700 thermocycler (Applied Biosystems, Foster City, CA, USA) or an I-Cycler (Biorad, Hercules, CA, USA) using the following conditions: an initial step at 95°C for 5 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing between 48 and 58°C for 90 s and extension at 72°C for 40 s, ending with a final extension of 10 min at 72°C. In those cases where specific amplification products within the expected size range were obtained, both strands were sequenced to verify the presence of nucleotide repeats. Sequencing reactions were carried out using the BigDye Terminator Cycle Sequencing Kit v1.1 (Applied Biosystems) and run on an automated DNA Sequencer (ABI 3100).

Of 34 microsatellite loci initially screened, 16 showed sufficient levels of polymorphism (three alleles or more) to be considered for further development as multiplexes in all six duiker species and sympatric artiodactyls *T. spekei* and *H. aquaticus*. The other 18 microsatellite loci were rejected because they were either monomorphic, did not amplify, or yielded non-specific products (Table 1). Twelve microsatellite loci were combined into three multiplex assemblages of four loci each. The loci within each multiplex were selected because they successfully co-amplified with one another and exhibited minor overlap in size. In cases where overlap was unavoidable, adjacent loci were labeled with different dye colors to minimize the risk of genotype error (Tables 2, 3). Co-amplification of microsatellite loci was carried out with the QIAGEN multiplex kit using the following conditions: an initial step at 95°C for 15 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C (multiplex 1 and 3) or 58°C (multiplex 2) for 90 s and extension at 72°C for 60 s, ending with a final extension step at 60°C for 45 min. This final step ensures the addition of the "A artifact" to the 3' end of the double-stranded DNA, making the scoring of genotypes more consistent across samples (Brownstein *et al.*, 1996).

Following multiplex amplification, 1.0µL of PCR product was combined with 0.8µL of 500bp HD500 Rox size standard (ABI), 11.2µL of formamide and run on an ABI 3100 capillary based sequencing machine. Allele fragment sizing was performed using GENEMAPPER v4.0 (ABI) and binning of raw data was carried out using the program FLEXIBIN (Amos *et al.*, 2007). Test of deviations from Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) for two species (*C. monticola*, *C. callipygus*) with sample sizes of 20 each were carried out using the program ARLEQUIN v3.11 (Schneider *et al.*, 2000). A Bonferroni correction was employed to determine the appropriate critical value for rejection of the null hypothesis for both HWE and LE tests assuming an alpha value of 0.05. The possibility of null alleles, large allele dropout and scoring error due to stuttering were also assessed using the program MICROCHECKER (Van Oosterhout *et al.*, 2004).

Table 1. Levels of polymorphism in the genus *Cephalophus* for 34 microsatellite loci originally isolated in other Bovids.

Locus	Origin: Genus (subfamily)	Primers (5'-3')	Tm	[MgCl2] (mM)	Genbank Acc. # (Species)	Amplification in *Cephalophus spp.*
BM121	*Bos* (Bovinae)[1]	F: tggcattgtgaaaagaagtaaa R: actagcactatctggcaagca	58-60	2.3	*GU265866 (C. monticola)* *GU265867 (C. nigrifrons)*	Polymorphic
BM1225	*Bos* (Bovinae)[1]	F: tttctcaacagaggtgtccac R: acccctatcaccatgctctg	54	2	*GU265874 (C. silvicultor)* *GU265875 (C. monticola)* *GU265876 (C. leucogaster)*	Polymorphic
BM143	*Bos* (Bovinae)[1]	F: acctgggaagcctccatatc R: ctgcaggcagattctttatcg	56-58	2	*GU265868 (C. silvicultor)*	Polymorphic
BM1862	*Bos* (Bovinae)[1]	F: aagcaaaaaggctgatggc R: ttgcagatactggcaagtgg	48-60	2	*GU265877,78 (C. nigrifrons)* *GU265879 (C. dorsalis)* *GU265881 (C. monticola)*	Polymorphic
BM2113	*Bos* (Bovinae)[1]	F: gctgccttctaccaaataccc R: cttcctgagagaagcaacacc	54	2	*GU265882 (C. monticola)* *GU265883 (C. leucogaster)*	Polymorphic
BM2830	*Bos* (Bovinae)[1]	F: aatgggcgtataaacacagatg R: tgagtcctgtcaccatcagc	50	2	*GU265884 (C. nigrifrons)*	Polymorphic
BM4513	*Bos* (Bovinae)[1]	F: gcgcaagtttcctcatgc R: tcagcaattcagtacatcaccc	54	2.1	*GU265885 (C. monticola)*	Polymorphic
BM848	*Bos* (Bovinae)[1]	F: tggttggaaggaaaacttgg R: cctctgctcctcaagacac	54	1.7	*GU265870 (C. silvicultor)* *GU265871 (C. monticola)*	Polymorphic
BM864	*Bos* (Bovinae)[1]	F: tggtagagcaatatgaaggcc R: ggaaatccaagaaagagggg	56-58	2	*GU265872 (C. monticola)* *GU265873 (C. nigrifrons)*	Polymorphic
BRRIBO	*Bos* (Bovinae)[1]	F: cacccgtaccctcactgc R: tcacaaccctcttctcaccc	58-60	2	*GU265886,87 (C. nigrifrons)* *GU265888 (C. leucogaster)*	Polymorphic
ETH225	*Bos* (Bovinae)[3]	F: gatcaccttgccactatttcct R: acatgacagccagctgctact	54	1.8	*GU265889 (C. monticola)* *GU265890 (C. leucogaster)*	Polymorphic
INRA005	*Bos* (Bovinae)[5]	F: caatctgcatgaagtataaatat R: cttcaggcataccctacacc	54	2	*GU265891 (C. nigrifrons)* *GU265892 (C. leucogaster)*	Polymorphic

Table 1 (continued). Levels of polymorphism in the genus *Cephalophus* for 34 microsatellite loci originally isolated in other Bovids.

INRA040	*Bos* (Bovinae)[5]	F: tcagtctccaggagagaaaac R: ctctgccctggggatgattg	54	2	*GU265893 (C. monticola)* *GU265894 (C. leucogaster)*	Polymorphic
MM12	*Capra* (Caprinae)[8]	F: caagacaggtgtttcaatct R: atcgactctggggatgatgt	58	1.9	*Too short to be reported*	Polymorphic
SR12	*Capra* (Caprinae)[11]*	F: tgaccaggtgactaacac R: aatctgatttcatttcatg R3: gactgagcgacttcactctc	54	1.7	*GU265895 (C. silvicultor)* *GU265896 (C. monticola)*	Polymorphic
CSSM66	*Bos* (Bovinae), Ovis (Caprinae)[2]	F: acacaaatcctttctgccagctga R: aatttaatgcactgaggagcttgg	-	-	-	Polymorphic
BM1706	*Bos* (Bovinae)[1]	F: acaggacggtttctccttatg R: cttgcagtttcccatacaagg	48-50	2	*GU265880 (C. monticola)*	Poor amplification
BM203	*Bos* (Bovinae)[1]	F: gggtgtgacattttgttccc R: ctgctcgccactagtccttc	56-58	2	*GU265869 (C. nigrifrons)*	Monomorphic
BM6121	*Bos* (Bovinae)[1]	F: ctgtttgctataattttgtggagg R: tggcattctacgagaccaca	-	-	-	Monomorphic
McMA49	*Ovis* (Caprinae)[7]	F: gtattctgttgtgtgctacagc R: gatgcaaattatatggatgttg	-	-	-	Monomorphic
SR24	*Capra* (Caprinae)[11]	F: agcaagaagtgtccactgacag R: tctaggtccatctgtgttattgc	-	-	-	Monomorphic
BM305	*Bos* (Bovinae)[1]	F: acacaataagagtgtggcatcc R: gtgtccttttgactcactgtgc	-	-	-	Nonspecific
BM757	*Bos* (Bovinae)[1]	F: tggaaacaatgtaaacctggg R: ttgagccaccaaggaacc	-	-	-	Nonspecific
OarFCB11	*Ovis* (Caprinae)[10]	F: gcaagcaggttctttaccactagcacc R: ggcctgaactcacaagttgatatatctatcac	-	-	-	Nonspecific

Table 1 (continued). Levels of polymorphism in the genus *Cephalophus* for 34 microsatellite loci originally isolated in other Bovids.

BMC701	*Bos* (Bovinae)[1]	F: tgatttccttttccagacttcc R: atgggttccagcacaatttt	-	-	-	No amplification
BM1818	*Bos* (Bovinae)[1]	F: agctgggaatataaccaaagg R: agtgctttcaaggtccatgc	-	-	-	No amplification
BM226	*Bos* (Bovinae)[1]	F: attgccttgtccgtgtatcc R: ccggctgaattgctataagc	-	-	-	No amplification
BM4505	*Bos* (Bovinae)[1]	F: ttatcttggcttctgggtgc R: atcttcacttgggatgcagg	-	-	-	No amplification
BM6017	*Bos* (Bovinae)[1]	F: tcttctgttttcctccatccc R: ggaaactagcttatgctgtggg	-	-	-	No amplification
BM720	*Bos* (Bovinae)[1]	F: acatctcattcttgtgtcatgg R: gaaattcagtttagggttcccc	-	-	-	No amplification
HEL1	*Bos* (Bovinae)[4]	F: caacagctatttaacaagga R: aggctacagtccatgggatt	-	-	-	No amplification
MAF70	*Ovis* (Caprinae)[6]	F: gcaggactctacggggcctttgc R: cacggagtcacaaagagtcagacc	-	-	-	No amplification
OarAE16	*Ovis* (Caprinae)[9]	F: ctttttaatggctcggtaatattcctc R: catcagaggaatgggtgaagacgtgg	-	-	-	No amplification
OarFCB266	*Ovis* (Caprinae)[10]	F: ggcttttccactacgaaatgtatcctcac R: caccacataccaaacacacagcctgc	-	-	-	No amplification

1. Bishop *et al.* (1994); 2. Barendse *et al*. (1994); 3. Steffen *et al.* (1993); 4. Kaukinen, Varvio (1993); 5. Vaiman *et al.* (1994); 6. Crawford *et al.* (1995); 7. Beh *et al.* (2000); 8. Mommenst *et al.* (1994); 9. Penty *et al.* (1993); 10. Buchanan, Crawford (1993); 11. Kogi *et al.* (1995).
*Reverse primer R3 was designed by authors and amplifies a smaller amplicon and is nested within F/R of SR12 (Kogi *et al.* 1995).

Table 2. Population genetic analysis of 12 dinucleotide microsatellite loci based on 20 individuals from two duiker species: *Cephalophus callipygus* and *C. monticola*

Locus	Fluorescent dye	Multiplex set	N_A	H_O	H_E	P_{HW}	Allele size range
C. callipygus							
BM2113	NED	1	9	0.7966	0.7895	0.7921	120-150
INRA040	HEX	1	10	0.7859	0.6000	0.0483	135-187
BM1225†*	NED	1	9	0.8436	0.4500	0.0000	221-237
BRRIBO [NT]	FAM	1	7	0.5808	0.5500	0.1317	236-250
BM143	FAM	2	8	0.7718	0.7500	0.8221	87-109
INRA005 [NT]	FAM	2	5	0.6092	0.4615	0.1088	129-153
BM1862 [NT]	HEX	2	5	0.3949	0.4500	1.0000	184-198
BM864	HEX	2	6	0.8681	0.5714	0.1594	226-248
MM12* [NT]	NED	3	3	0.4718	0.2000	0.0000	77-85
BM121 [NT]	FAM	3	5	0.4256	0.3500	0.4020	110-120
BM848 [NT]	HEX	3	3	0.5539	0.7000	0.0175	199-213
SR12 *[NT]	FAM	3	3	0.6269	0.4500	0.0007	217-231
C. monticola							
BM2113	NED	1	10	0.8949	0.8500	0.4694	122-144
INRA040*†	HEX	1	14	0.9218	0.6500	0.0034	158-184
BM1225	NED	1	9	0.7653	0.6842	0.5140	217-235
BRRIBO*†	FAM	1	12	0.9093	0.2500	0.0000	231-257
BM143*†	FAM	2	10	0.8782	0.2500	0.0000	87-109
INRA005	FAM	2	8	0.7701	0.5294	0.0212	133-189
BM1862	HEX	2	7	0.7821	1.0000	0.0069	187-204
BM864*†	HEX	2	5	0.6934	0.4118	0.0000	226-256
MM12 [NT]	NED	3	4	0.1462	0.1000	0.0716	85-91
BM121*†	FAM	3	14	0.9192	0.5000	0.0000	104-158
BM848	HEX	3	15	0.9141	0.7000	0.0188	190-230
SR12	FAM	3	10	0.8777	0.7368	0.0045	207-245

NT, not tested in MICROCHECKER because the locus had more than 50% of the alleles within one allele class

N_A, number of alleles detected per locus; H_O, observed heterozygosity; H_E, expected heterozygosity; P_{HW}, probability of Hardy–Weinberg equilibrium test

†Null alleles detected in MICROCHECKER.

*Deviated from HWE after Bonferroni correction.

Table 3. Cross-species amplification of twelve dinucleotide microsatellite loci among six individuals of the following species: *Cephalophus silvicultor, C. nigrifrons, C. dorsalis*, *C. leucogaster, Tragelaphus spekei* and *Hyemoschus aquaticus*. Locus name, Fluorescent dye, multiplex set, size range and, between parentheses, number of alleles per locus, are shown.

Locus	Fluorescent dye	Multiplex set	*C. silvicultor*	*C. nigrifrons*	*C. dorsalis*	*C. leucogaster*	*T. spekei*	*H. aquaticus*
BM2113	NED	1	124-134 (4)	115-133 (4)	124-138 (4)	115-123 (2)	114-142 (5)	106-126 (4)
INRA040	HEX	1	147-195 (6)	187-195 (5)	156-176 (7)	147-151 (3)	134-196 (7)	139-171 (5)
BM1225	NED	1	213-219 (4)	213-225 (6)	214-216 (2)	223-227 (3)	227-265 (7)	215-259 (3)
BRRIBO	FAM	1	223-257 (8)	240-260 (9)	235-259 (7)	236-251 (7)	225-249 (6)	241-255 (5)
BM143	FAM	2	89-113 (6)	92-108 (5)	83-107 (7)	93-113 (5)	79-105 (4)	82-100 (5)
INRA005	FAM	2	135-155 (4)	154-177 (7)	128-157 (5)	154-170 (6)	144-170 (4)	114-144 (4)
BM1862	HEX	2	186-212 (6)	187-202 (5)	196-198 (2)	190-202 (4)	187-199 (7)	196-200 (4)
BM864	HEX	2	226-250 (5)	226-250 (4)	226-248 (3)	224-248 (4)	226-238 (5)	226-288 (5)
MM12	NED	3	75-79 (3)	79-85 (4)	77 (1)	81 (1)	81-85 (2)	77-81 (2)
BM121	FAM	3	120 (1)	110-122 (2)	110-160 (7)	110 (1)	118-122 (3)	104-134 (6)
BM848	HEX	3	197-215 (4)	195-201 (4)	194-224 (7)	-	199-219 (6)	204-224 (4)
SR12	FAM	3	233-237 (3)	223-253 (7)	209-239 (8)	229-239 (5)	211-229 (5)	228-234 (3)

4.3. Results and Discussion

In *C. callipygus* and *C. monticola,* the number of alleles ranged from three to ten (mean=6.08, SD=2.5) and four to fifteen (mean=9.83, SD=3.51) respectively (Table 2). Average expected heterozygosity was 0.64 (SD=0.17) in *C. callipygus* and 0.79 (SD=0.22) in *C. monticola*. In *C. callipygus*, three loci (BM1225, MM12 and SR12) deviated from HWE after Bonferroni correction. Deviations at one of these loci (BM1225) were due to the presence of null alleles (Table 2). In *C. monticola*, five loci (BRRIBO, BM143, BM864, BM121 and INRA040) were out of HWE, all of which appeared to be due to null alleles (Table 2). There was no evidence of either large allele dropout or scoring error due to stuttering in either species. A recent analysis of these two species (sampled from the same population as the current study) using mitochondrial DNA sequences revealed higher than expected nucleotide diversity, suggesting some population structure may exist over small distances (Eaton *et al.*, 2009). If this is the case, heterozygote deficiencies at some loci could also be due to a Wahlund effect. To discriminate between these two possible factors, it is recommended to first sample from a known panmictic population and, if deviations persist, to redesign one or both primers from loci found to be consistently out of HWE. If null alleles cannot be eliminated, then heterozygote deficiencies can be rectified using methods implemented in MICROCHECKER (Van Oosterhout *et al.*, 2004).

Initial analysis with ARLEQUIN (Schneider *et al.* 2000) indicated that two pairs of loci (BM848 and SR12, P=0.018; INRA005 and BM864, P=0.029) deviated significantly from LE for *C. callipygus*. However, after Bonferroni correction, neither locus was significantly linked. In the case of *C. monticola,* none of the twelve loci deviated significantly from LE (Table 2). All twelve microsatellite loci amplified across the remaining six other species, except BM848 which failed in *C. leucogaster* (Table 3).

An additional four loci (BM4513, BM2830, ETH225 and CSSM66, see Table 1) not included in the three multiplex combinations tested herein could also be used individually and/or combined in a different multiplex set across the same set of taxa. These loci were not included within the original three multiplexes because they either overlapped in size range and fluorescent dye or yielded poor amplification success when combined with the other candidate loci.

All twelve microsatellite loci assessed in the present study proved to be highly polymorphic with substantial allelic richness across all tested artiodactyl species. These optimized multiplexes will benefit future studies of central African artiodactyls and offer considerable savings in both time and cost. Applications of these microsatellite multiplexes evaluated here include but are not limited to studies of individual identification, parentage analysis, population size estimation and fine-scale analyses of population genetic structure.

Chapter 5: testing Pleistocene refugia and riverine barrier hypotheses of diversification using central African duikers as a model (genera *Cephalophus* and *Philantomba*)

5.1. Introduction

The tropics are known to be areas of high species richness and endemism (Brooks *et al.*, 2006; Da Silva *et al.*, 2005; Joseph *et al.*, 1995; Linder, 2001; Mittermeier *et al.*, 1998; Myers *et al.*, 2000; Reid, 1998) where several hypotheses have been proposed in order to explain evolutionary processes underlying species diversification. Although, most studies have been carried out in South American (Brumfield, Capparella, 1996; Da Silva *et al.*, 2005; Hayes, Sewlal, 2004; Patton *et al.*, 1994) and Australian (Joseph *et al.*, 1995; Schneider, Moritz, 1999) rainforests while little attention has been paid to central Africa (Anthony *et al.*, 2007; Fjeldsa, Lovett, 1997; Maley, 1996a; Muloko-Ntoutoume *et al.*, 2000). Suggested hypotheses of tropical diversification were mostly based on vicariance (Ayres, Clutton-Brock, 1992; Bush, 1975; Bush, 1994; Cracraft, Prum, 1988; Haffer, 1969; Haffer, 1997; Wallace, 1852), ecological (Endler, 1977; Endler, 1982), and historical (Haffer, 1969) factors which may have affected studied organisms. However, the African tropic was differently impacted by the last glacial maximum of the Pleistocene with a much colder and drier climate (DeBusk, 1998) that might have favored the formation of forest refuges, as opposed to the Amazon where it was much wetter (Anhuf *et al.*, 2006; Cardenas *et al.*, 2011). Nevertheless, regardless of the tropical region considered, two hypotheses have attracted a lot of attention: the Pleistocene forest refugia (Haffer, 1969) and the riverine barrier hypothesis (Wallace, 1852).

The Pleistocene refugia hypothesis was initially proposed by Haffer (1969) to explain patterns of species richness in the Amazon basin. This hypothesis states that during the drier and cooler phases of the Pleistocene, tropical forests contracted into smaller fragments or refugia which in turn lead to the isolation and subsequent speciation of forest-associated taxa (Haffer, 1969; Haffer, 1997). These forest refugia may have occurred in upland areas (Bakker, Coetzee, 1988; Bonnefille *et al.*, 1990; DeBusk, 1998) or river basins (lowland refugia) where the higher moisture content or presence of major river bodies (Colyn *et al.*, 1991; Sangen, 2011; Sangen *et al.*, 2011) allowed forest vegetation to persist. Researchers generally recognize three major predictions of the Pleistocene refugia hypothesis: (1) fragmentation of formerly contiguous forest in isolated blocks (forest refugia) during glacial maxima that lead to allopatric fragmentation and divergence between isolated populations, (2) expansion of populations during warmer interglacial periods to fill areas of suitable habitat leaving characteristic genetic signatures of population expansion, and (3) occurrence of regions of secondary contact where genetic admixture between expanding neighboring refugial populations may have occurred (Haffer, 1969; Hewitt, 1996; Lessa *et al.*, 2003).

In contrast, proponents of the riverine barrier hypothesis (Wallace, 1852) argue that rivers have constituted important barriers to gene flow and if maintained over sufficient time may have led to divergence and ultimately speciation (Bush *et al.*, 1992; Haberle, 2005; Kershaw *et al.*, 2007; Kirkpatrick, Fowler, 1998). The three predictions of the riverine barrier hypothesis are: (1) sister species and/or populations should be separated by broad rivers, (2) cross-bank genetic differentiation should increase from headwaters to the mouth, (3) species and/or populations that ordinarily do not cross water bodies should show greater genetic differentiation pattern than those that are able to do so (Haffer, 1997; Hayes, Sewlal, 2004; Patton *et al.*, 1994).

A handful of studies to date have tested the role of rivers and refugia in evolutionary diversification of central African taxa. Anthony and colleagues (2007) assessed the role of Pleistocene forest refugia and rivers in shaping pattern of genetic diversity in central African gorillas (*Gorilla gorilla*). They found evidence of forest refugia in upland areas of Gabon (Massif du Chaillu, Mont Doudou and Monts de Cristal) and Equatorial Guinea (Monte Alén) and a putative lowland refugium in Central African Republic. The same study also showed that several rivers (Sangha, Ogooué, and Ivindo/Ayina Rivers) may constitute barriers to gorilla gene flow. Similarly, previous studies of the rainforest tree *Acoumea klaineana* (Born *et al.*, 2011; Muloko-Ntoutoume *et al.*, 2000) suggested a refugial effect of Massif du Chaillu, Mont Doudou, Monts de Cristal, and Massif du Mayombe in Gabon, while studies of mandrill (*Mandrillus sphinx*) populations in Gabon (Telfer *et al.*, 2003) suggested a role for the Ogooué River in structuring genetic diversity in these primates. In addition, Eriksson and colleagues (2004) assessed the role of several rivers in shaping the genetic diversity of five bonobo (*Pan paniscus*) populations on the southern bank of the Congo River in the Democratic Republic of the Congo (DRC). Genetic distance was only significantly correlated with geographic distance when distance measures were made around the river course. In contrast, there was no evidence of an isolation by distance association based on straight line distances between riverbanks (Eriksson *et al.*, 2004). Findings from a phylogeographic study of the murid rodent *Praomys misonnei* across West, Central and East Africa also suggested that several rivers and refugia may have played an important role in shaping pattern of genetic diversity in this rodent (Nicolas *et al.*, 2011). The authors also identified several putative Pleistocene refugia and putative riverine barriers across their study area (Nicolas *et al.*, 2011). However all of these phylogeographic studies to date have been limited to single species studies and did not necessarily address the same study area. A comparative multi-species approach that incorporates multiple refugia and riverine barriers is a necessary next step in assessing the role of refugia and rivers in shaping species diversity (Bermingham, Moritz, 1998; Joseph *et al.*, 1995).

Forest duikers (subfamily Cephalophinae) constitute an ideal group for testing the forest refugia and riverine barrier hypotheses. Firstly, these species are believed to have recently originated in the Pleistocene (Vrba, 1995), making them more likely to have been heavily influenced by changes in forest cover and river bodies during that time. Secondly, duikers favor forested environments whose distribution is likely to have been affected by past climate change (Bush *et al.*, 1992; Haberle, 2005; Kershaw *et al.*, 2007; Kirkpatrick, Fowler, 1998). Thirdly, multiple species within this group occur in sympatry across central Africa (Dubost, 1984; Lahm, Tezi, 2006; Newing, 2001), allowing us to evaluate whether all species genetic patterns reflect a common biogeographical history (Bermingham, Moritz, 1998; Joseph *et al.*, 1995). Fourthly, their geographic distribution spans the river banks of the Ogooué and Sanaga, both of which are thought to be important biogeographical barriers (Anthony *et al.*, 2007; Muloko-Ntoutoume *et al.*, 2000; Nicolas *et al.*, 2008; Nicolas *et al.*, 2006; Telfer *et al.*, 2003). Lastly, they are amenable to non-invasive genetic and suitable storage methods and extraction protocols are available (Idaghdour *et al.*, 2003; Murphy *et al.*, 2000; Nsubuga *et al.*, 2004; Soto-Calderon *et al.*, 2009; Wasser *et al.*, 1997).

The main objectives of the present study are to use mitochondrial and nuclear microsatellite data obtained from geo-referenced duiker feces to assess both historical and contemporary population genetic structure and evaluate the role that riverine barriers and Pleistocene refugia may have played in their evolutionary diversification. Due to the need to gather adequate sample sizes for population genetic analyses, this study focused on the three most commonly encountered central African forest antelope in our sampling area: the bay duiker

(*Cephalophus dorsalis*), the Peter's duiker (*Cephalophus callipygus*), and the blue duiker (*Philantomba* monticola). The specific questions which are the focus of the present study are:

(1) Does analysis of genetic structure identify differentiated populations of which geographic distribution coincide with suggested Pleistocene refugia?
(2) Does analysis of genetic structure identify differentiated populations of which geographic distribution coincide with opposed riverbanks of the Ogooué and Sanaga?
(3) Is there any signature of population expansion out of suggested Pleistocene refugia?
(4) Are areas of genetic admixture geographically located between suggested Pleistocene refugia?
(5) Does genetic differentiation increases along the Ogooué and Sanaga rivers from the source to the mouth?
(6) Do all three tested duiker species harbor similar pattern of genetic structure and biogeographic history?

5.2. Methods

5.2.1. Sample collection and DNA extraction

A total of 2074 samples (2040 feces + 34 museum and zoo tissues) were collected from 46 sites across 9 countries in Africa (See Figure 1 and Supplementary Table 1). Collection sites included the candidate upland refugia of (1) Monte Mitra and Monte Alén National Park (Site 16) in Equatorial Guinea, (2) Monts de Cristal National Park (Site 17) and (3) Massif du Chaillu (Site 33) in Gabon, (4) highlands of southeast Nigeria and Cameroon (Sites 2, 3), and (5) the putative lowland refugia of Campo Ma'an area in southwestern Cameroon (14) and the Sangha basin (Sites 37, 38, 39) at the border between the Central African Republic, Cameroon, and the Republic of Congo. Fecal samples were also collected on both banks of multiple sites along the two potential riverine barriers Ogooué (Sites 20, 21, 22, 23, 24, 25, 26, 29, 30, 31, 32) and Sanaga (Sites 5, 6, 7, 8, 9, 10, 11, 12, 13) (Figure 1). In order to collect samples from as much representative habitat types as possible 17 additional sampling sites were included (see Table 1). At each site, fresh fecal pellets (< 24 hours) were sampled opportunistically following a general compass bearing along the path of least resistance using the REConnaissanCE (RECCE) method (Walsh, White, 1999).

One to three pellets were either placed into a vial containing 5 g of silica gel beads (Sigma) or a 2 mL cryovial (Fisher) containing 1.2 mL of RNA*later* (Ambion, Inc.), since previous studies have shown that these two storage methods maximize nuclear DNA extraction yields (Soto-Calderon *et al.*, 2009). For each sample, the location, GPS coordinates, and major habitat type in which the sample was collected were recorded. DNA extraction was carried out using the QIAGEN DNA stool Minikit (Qiagen, Valancia, CA). Museum and zoo samples of known geographic origin were also obtained from sites in Diecke (Mount Nimba) (Site 45) in the Republic of Guinea; Cape province (Site 44) in South Africa; Dja (Site 15) and Bamenda (Site 4) in Cameroon; Reserve de la Lefini (42), Brazzaville area (43), and Parc National d'Odzala (Site 36) in the Republic of the Congo; Malounga (35) in Gabon; Parc National de Ngotto (Site 39) in Central African Republic; and Kisangani (41) in the Democratic Republic of the Congo (Supplementary Table 1). These samples were either extracted from teeth or from muscle preserved in ethanol as previously described (Ntie *et al.*, 2010a).

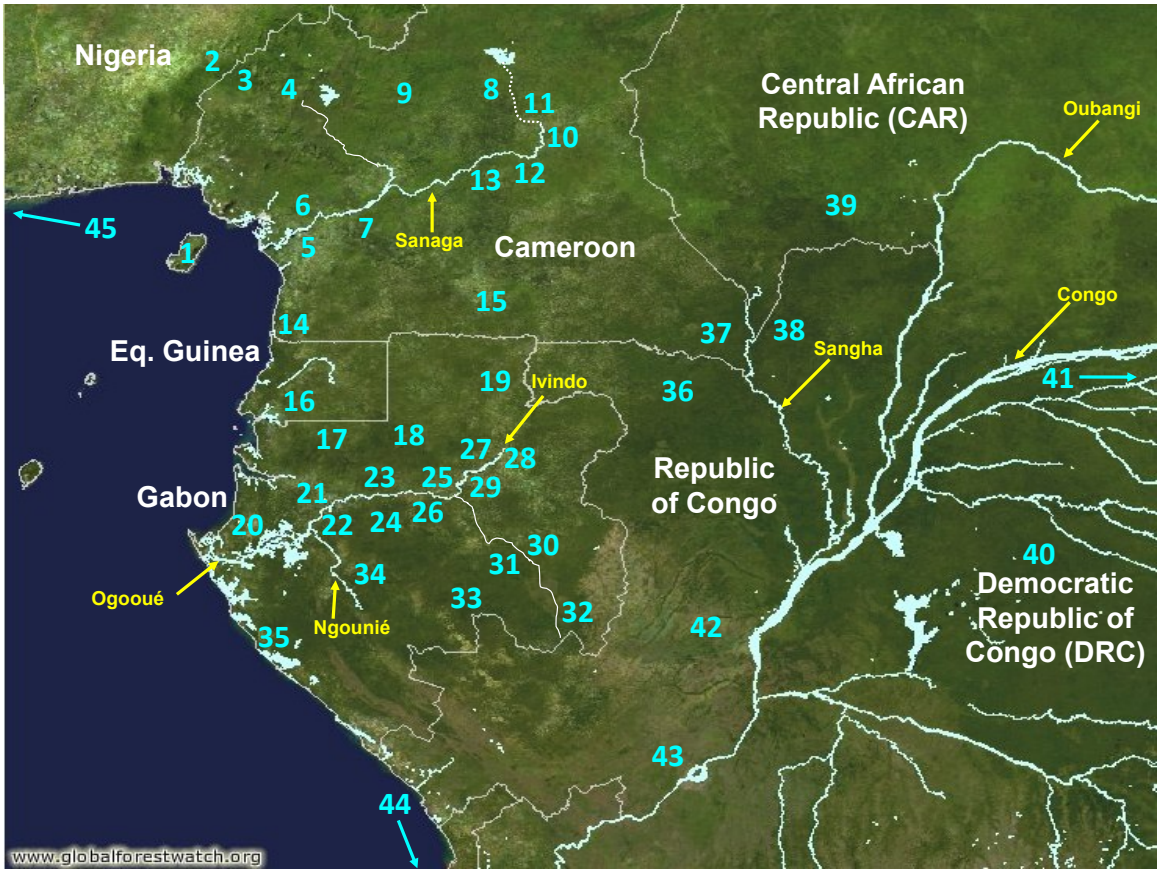


Figure 1. Sites where fecal samples have been collected and identified to species level for the three most frequently encountered species *C. dorsalis, C. callipygus, and P. monticola*. Major rivers are indicated in yellow. See Supplementary Table 1 for site names.

5.2.2. DNA amplification, sample identification, and microsatellite genotyping

An approximately 650 base pairs (bp) long fragment of the mitochondrial control region was amplified and sequenced following previously optimized conditions (Ntie *et al.*, 2010a). Unknown fecal DNA sequences were aligned with reference tissue DNA sequences of known species identity and used to construct a phylogenetic neighbor-joining tree (Ntie *et al.*, 2010a). The species identity of all fecal samples was then recovered by observing the species-specific clade in which unknown fecal samples fell. A quantitative PCR assay (Soto-Calderon *et al.*, 2009) based on a 96 bp of the vertebrate p53 tumor suppressor gene (Bellis *et al.*, 2003) was also used to assess nuclear DNA quantity and determine the number of PCR replicates required to obtain a reliable genotype (see Morin *et al.,* 2001). Hence, duiker fecal samples with 25-49.9 pg/µl of nuclear DNA or higher were genotyped a minimum of four times and those of 50 pg/uL of DNA or more were typed a minimum of three times as previously determined (see Soto-Calderon *et al.,* 2009). Selected samples were genotyped with 12 polymorphic microsatellites assembled into three multiplex reactions of four loci each following previously optimized conditions (Ntie *et al.*, 2010b).

5.2.3. Analysis of mitochondrial Control region sequences

5.2.3.1. Identification of mitochondrial haplogroups

A neighbor-joining phylogenetic analysis was conducted on the mitochondrial control region of *C. dorsalis* (598 bp), *C. callipygus* (595 bp), and *P. monticola* (757 bp) respectively from

113, 457, and 359 individuals. A neighbor-joining method was chosen because: (1) it is very efficient at producing a single topology that minimizes the total branch length of a tree and is very close to the optimal tree (Saitou, Nei, 1987), (2) it is consistent with many models of nucleotide evolution, and (3) it can be used with very large datasets. This analysis was carried out to identify major phylogenetic lineages that may correspond to putative refugia and differentiated populations on opposite banks of the Sanaga and Ogooué rivers. The best model of nucleotide substitution was assessed with the program jModelTest 0.1 (Posada, Crandall, 1998) and selected using the Akaike delta value, which is the minimum theoretical information criterion (AIC) (Akaike, 1974). The *C. dorsalis* tree was based on the HKY model (Hasegawa *et al.*, 1985) with the proportion of invariant sites allowed and a gamma model of among site variation (alpha = 0.53). The *C. callipygus* tree was based on the General Time-Reversible (GTR) model (Tavaré, 1986) with the proportion of invariant sites allowed and a gamma model of among site variation (alpha = 0.5). The *P. monticola* tree was based on the TrN model (Tamura, Nei, 1993) with a gamma model of among site variation (alpha = 0.31). Prior to phylogenetic analysis, mtDNA control region sequences were collapsed within each sampling site to 96 unique haplotypes for *C. dorsalis*, 146 for *C. callipygus* (sequences differing at 20 nucleotide positions and less), and 121 for *P. monticola* (sequences differing at 23 nucleotide positions and less) using the program Collapse v1.2 (available from http://darwin.uvigo.es). For ease of graphical illustration, the number of nucleotides considered for collapsing was determined empirically in order to end up with less than 150 haplotypes total per species. A neighbor-joining bootstrap consensus tree of 1000 replications of these data was carried out using PAUP v4.0 beta (Swofford, 2003). Haplogroups were defined as reciprocally monophyletic clades with 50% or more bootstrap support (see Supplementary Figures 1, 2, and 3).

The program NETWORK (version 4.6.1.0) was used to create an intraspecific Minimum Spanning Network (MSN) (Bandelt *et al.*, 1999) to better visualize the pattern of evolutionary relationships in haplotypes sampled from putative refugia and opposing riverbanks. For practical reasons and to simplify data interpretation, mtDNA control region sequences were collapsed into 16 major haplotypes for *C. dorsalis* (sequences differing at 17 nucleotide positions and less), 27 haplotypes for *C. callipygus* (sequences differing at 27 nucleotide positions and less), and 25 *P. monticola* (sequences differing at 31 nucleotide positions and less) using the program Collapse v1.2 (available from http://darwin.uvigo.es) (see Supplementary Figures 4, 5, and 6)

5.2.3.2. Spatial and non spatial Analyses of Molecular Variance (SAMOVA and AMOVA)

SAMOVA 1.0 (Dupanloup *et al.*, 2002) was used to define the structure that maximizes the among groups variance component while taking into account the geographic proximity of sample sites. The goal here is to identify genetic structure without prior information of the population groupings among putative refugia and riverbanks. ARLEQUIN 3.5.1.3 (Excoffier *et al.*, 2005) was used to conduct an AMOVA in order to test the significance of inferred SAMOVA groupings and test hypotheses of genetic structure according to putative refugia and riverbanks. SAMOVA and AMOVA were performed on populations or groups of populations with a sample size of four or more individuals. Because of low sample sizes in *C. dorsalis*, a few sites with similar haplotype composition and geographic position were grouped together.

For all three species (*C. dorsalis*, *C. callipygus*, and *P. monticola*), the dataset was first subdivided into two groups to reflect major hypothesized refugia; one group comprised

samples from each putative upland, lowland, and/or fluvial refugium respectively while the other was made up of samples from all other sites sampled across central Africa (respectively Tables 2, 3, and 4). We also tested the hypothesis that the dataset was genetically structured by multiple refugia (Tables 2, 3, and 4). We also tested the barrier effect of the Sanaga and/or Ogooué by subdividing the dataset into two groups north and south of each river (Tables 2, 3, and 4). To test the combined effects of both the Sanaga and Ogooué river bodies, we structured the entire dataset into four groups that was made up of samples from the north and south of the Sanaga, and north and south of the Ogooué (Tables 2, 3, and 4). In the case of *P. monticola*, Nigerian samples were excluded from all AMOVA analyses as the genetic differentiation of this group was evident based on the phylogenetic analysis (see below). In the case of *C. callipygus*, AMOVA could not be tested specifically on populations flanking the Sanaga River due to the small sample sizes.

5.2.3.3. Genetic diversity and demographic history

The program DnaSP V5.1 (Librado, Rozas, 2009) was used to calculate haplotype diversity (Hd) of individual haplogroups (Nei, Tajima, 1983), the average number of nucleotide differences per site between sequences (Pi) (Jukes, Cantor, 1969), and the mean number of mutations per nucleotide per sequence (Theta) (Watterson, 1975). The latter tests are critical for identifying sites with high genetic diversity such as hypothesized refugia or zones of secondary contacts between different refugial populations. Demographic history was assessed using Tajima's D, Fu and Li's D* and F* neutrality tests (no mutation, no selection, no migration, no genetic drift, random mating, and large population size). Fu's Fs and R^2 were utilized because they are the most powerful tests for detecting population growth (Fu, 1997; Ramos-Onsins, Rozas, 2002). In fact, R^2 is better suited for population sample size of 10 to 20, while Fu's Fs statistics is best for larger sample size of 20 or more (Ramos-Onsins and Rozas, 2002). All tests of deviation from neutrality were carried out within haplogroups defined by the population-level of each species phylogeny (see supplementary Figures 1, 2, and 3). Haplogroups with sample sizes of less than 4 samples were not tested because of software requirements (see DnaSP V5.1 manual).

5.2.4. Analysis of nuclear microsatellites

5.2.4.1. Expected heterozygosity and test of Linkage Equilibrium (LE) and Hardy-Weinberg Equilibrium (HWE)

ARLEQUIN 3.5.1.3 (Excoffier *et al.,* 2005) was used to calculate the expected heterozygosity, and deviations from LE and HWE for all 11 polymorphic microsatellite loci within each population. The significance of LE and HWE tests was assessed using an extension of Fisher's exact probability test with contingency tables (Slatkin, 1994) and a test analogous to Fishers's exact test on a two-by-two contingency table but extended to a triangular contingency table of arbitrary size respectively (Guo, Thompson, 1992). Expected heterozygosity is critical because it provides an estimation of the fraction of all individuals in a population who would be heterozygous at a given locus.. Tests of LE and HWE were performed in order to identify loci which are not randomly associated with one another and identify populations that do not conform to assumptions of an ideal population (no mutation, no selection, no migration, no genetic drift, random mating, and large population size). Expected heterozygosity and tests of LE, and HWE were performed on populations or groups of populations with sample sizes of four or more individuals.

5.2.4.2. Probability of Identity (P_{ID}) and relatedness (r) among samples

Gimlet version 1.3.3 (Valiere, 2002) was used to assess the minimum number of typed loci required in order to differentiate first order relatives ($P_{ID} \leq 0.05$). This information was critical because preliminary results showed that first order relatives could influence the outcome of the analyses. The minimum number of typed loci needed to differentiate first order relatives was calculated by multiplying across loci with the highest P_{ID} values until a minimum P-value of 0.05 or lower was attained. KINGROUP v2 (Konovalov *et al.*, 2004) was used to identify identical genotypes (r=1) and first order relative relationships ($r \geq 0.5$) between two or more individuals sampled from the same population. This information was retained in order to test the effect of first order relatives on nuclear genetic structure (see below). Once identified, samples with a duplicate genotype within each population were removed from downstream analyses.

5.2.4.3. Population genetic structure with the program STRUCTURE

Patterns of genetic structure were investigated using the Bayesian program STRUCTURE (Pritchard *et al.*, 2000). One advantage of STRUCTURE is that it uses multi-locus genotype data to assign individuals to clusters without *a priori* knowledge of their composition. Another advantage of STRUCTURE is the detection of admixed individuals from different genetic clusters which may indicate hybridization between refugial populations. The total number of steps in the Markov Chain Monte Carlo (MCMC) and burn-in were set respectively to one million and 100,000 generations with 10 separate simulations for each value of K. The admixture model was used and allele frequencies among populations were assumed to be correlated. Clustering of K populations within species was used to assess whether population differentiation was structured by putative refugia and/or riverbanks. In order to assess the effect of including first order relatives in all three species datasets, two independent STRUCTURE analyses were carried out with and without first order relatives. Additional STRUCTURE analyses were run on the *P. monticola* dataset to explore the effects of removing the highest order structure from the dataset since this can sometimes obscure more subtle population structure within derived K groupings.

Structure Harvester (Earl, Vonholdt, 2012) was used to determine the most probable value of K using the Delta K statistics procedure (Evanno *et al.*, 2005). Clumpp1.1.1 (Jakobsson, Rosenberg, 2007) was used to align all multiple runs for each K (because of label switching) and Distruct 1.1 (Rosenberg et al., 2003) was used to make graphical representations of STRUCTURE outputs.

5.2.4.4. Population genetic structure with the program GENELAND

The number of K clusters in each species' dataset and assignment of individual to each of these clusters were also inferred using the Bayesian clustering algorithm GENELAND 4.0.0 (Guillot *et al.*, 2005). The advantage of GENELAND over STRUCTURE is that it performs spatial genetic analysis of genetic data using both geographic and genetic information. In order to determine the most probable K, 10 million MCMC iterations with a thinning interval of 1,000 steps were performed for values of K varying between one and 10. Ten independent simulations were performed to check that replicate runs gave approximately the same parameter estimates. In order to take into account the effect of first order relatives on genetic structure, the maximum rate of the Poisson process (Lantuéjoul, 2002) was set to different values that reflected the size of the dataset with and without first order relatives, respectively:

65 and 54 for *C. dorsalis*; 310 and 233 for *C. callipygus*; 332 and 250 for *P. monticola*. The Poisson process (Lantuéjoul, 2002) is the statistical model implemented in GENELAND to infer and draw clusters from the provided genetic and geographic data. GENELAND analyses were also run on the *P. monticola* dataset with the NGO (Site 39) samples removed in order to explore the effects of removing the highest order structure. In this case, the maximum rate of the Poisson process was set to 221 for the dataset including individuals with first order relatives, 195 without. The maximum number of nuclei in the Poisson-Voronoi tessellation was set to different values among species: 195 and 162 for *C. dorsalis*; 930 and 699 for *C. callipygus*; 996, 750 for *P. monticola*. These latter numbers correspond to the value of the Poisson process multiplied by three, as advised by GENELAND manual. Similarly, the maximum number of nuclei in the Poisson-Voronoi tessellation for *P. monticola* dataset with and without the highest order structure was set to 663 and 585 respectively. Spatial uncertainty was set at ~2 km. The number of clusters was inferred from the modal value of K across ten runs. Runs were then sorted according to mean posterior density and only the best run was post-processed to obtain posterior probabilities of population membership for each individual and each pixel of the spatial domain using a burn-in of 100,000 iterations. Pixel number was set at 100 along both the X and Y axes. Similarly to STRUCTURE analysis, the best K may constitute differentiated populations from putative Pleistocene refugia and/or riverbanks and geographical boundaries between illustrated clusters may indicate admixture zones between refugial populations.

5.2.4.5. Test of Isolation by Distance using Isolation By Distance Web Service (IBDWS)

Isolation by distance test was performed using IBDWS v3.23 (available from http://ibdws.sdsu.edu/~ibdws/) (Jensen *et al.*, 2005) with 10,000 permutations to assess the statistical significance of the correlation between Slatkin's linearized measure of Fst, Fst/(1 – Fst), (Rousset, 1997)) and geographic distances. Slatkin's linearized Fst was used because it is more reliable and its correlation with geographic distance is easier to interpret when studying natural populations (Rousset, 1997). Isolation by distance test was calculated with both raw geographic distances and log transformed because differentiation between populations at larger distances follows the one-dimensional model and at smaller distance follows the two-dimensional model respectively (Rousset, 1997).

5.3. Results

5.3.1. Sample Identification

A total of 1146 fecal samples were identified to species level, of which three species appeared particularly well represented in our field collections and were subsequently chosen as targets for the present study. These species are: the bay duiker *C. dorsalis* (131 samples), Peter's duiker *C. callipygus* (475 samples), and blue duiker *P. monticola* (370 samples) (see Table 1). In addition, five other species were identified: the yellow-backed duiker *C. silvicultor* (39 samples), the black-fronted duiker *C. nigrifrons* (5 samples), the white-bellied duiker *C. leucogaster* (16 samples), the water chevrotain *Hyemoschus aquaticus* (71 samples), and the sitatunga *Tragelaphus spekeii* (39 samples). The remaining 928 fecal samples were not identified to species level because they failed to amplify.

Table 1. Sampling locations by country, site name, site numbers, letter code, and sample size for identified species in our sampling area (see Figure 1).

				Sample size							
Country	Sites names	Site number	Letter code	*C. dorsalis*	*C. callipygus*	*P. monticola*	*C. silvicultor*	*C. nigrifrons*	*C. leucogaster*	*H. aquaticus*	*T. spekei*
Cameroon	Campo Reserve	14	CPO	2	12	21					
Cameroon	Deng deng	10	DENG			3					
Cameroon	Bamenda	4	DIV			1					
Cameroon	Dja Reserve	15	Dja	3							
Cameroon	Ebo Forest	6	Ebo	9	1	19				1	
Cameroon	Douala-Edea Forest Reserve/Ekoth	5	Edea			8					2
Cameroon	Ekanga/Masseng/Biwali /Ngamba Enduum	13	EKA			3					
Cameroon	Kombe	7	KOM	2	1	5					
Cameroon	Parc National de Lobéké	37	LBK	5	16	4	1				
Cameroon	Linté/Ngambe Tikar	9	LIN	1			1				
Cameroon	Parc National de Mbam et Djerem/Wouchaba	8	MBJ	4	5					1	
Cameroon	Nkolomaken	12	NKO	5		6					
Cameroon	Lom/Pangar	11	PAN	2		1					
Cameroon	Takamanda Forest Reserve	3	TAK	6	2	10					
Central African Republic	Parc National de Ngotto	39	NGO	4		15					
Democratic Republic of Congo	Kisangani	41	KIS			1					
Democratic Republic of Congo	Parc National de Salonga	40	SA	1	5	1	3				
Equatorial guinea	Reserva Científica de la Caldera de Luba, Bioko	1	BKO		5	2		1			

Table 1 (continued). Sampling locations by country, site name, site numbers, letter code, and sample size for identified species in our sampling area (see Figure 1).

Equatorial guinea	Monte Mitra/Monte Alen	16	MM/MTA	6	1	35					
Gabon	Gamba, Iguela, Malounga	35	GA		2	33	3			15	
Gabon	Boumango, Parc National des Plateaux Batéké, Leconi, Bakoumba, Ossélé/Kessala	32	HAO	3	15	7	2		4	9	1
Gabon	Ivindo/Ipassa/Dji dji (West and East)	27, 28	IV	8	29	16	3		1	4	
Gabon	Langoué	29	LA	9	21	6	1	1	4		
Gabon	Parc National de la Lopé (north and south)	23, 24	LO	1	143	40	10		1	18	1
Gabon	Parc National des Monts de Cristal	17	MCR	12	1	16				1	
Gabon	Massif du Chaillu	33	MFCH	5	8	9			1		
Gabon	Parc National Minkébé	19	MKB	3	11	4	2		1		
Gabon	Moyen Ogooué (north and south)	21, 22	MOO			6	6			12	23
Gabon	Mitzic	18	MTZ		1				1		
Gabon	Ogooué Ivindo (north and south)	25, 26	OIV	9	54	7	1			5	
Gabon	Ogooué Lolo (north and south)	30, 31	OLO	6	46	11	3		1	3	
Gabon	Ogooué Maritime	20	OMA	1		1					12
Gabon	Parc National de Waka	34	WA		11	1					
Nigeria	Mbe, Oban, and Afi	2	NIG		6	35					
Republic of Congo	Brazzaville area	43	BRA		1						
Republic of Congo	Reserve de la Lefini	42	LEF	2		1					

Table 1 (continued). Sampling locations by country, site name, site numbers, letter code, and sample size for identified species in our sampling area (see Figure 1).

Republic of Congo	Parc National de Nouabelé Ndoki	38	ND/CO	1	60	30	3	3	2	2	
Republic of Congo	Parc National d'Odzala	36	ODZ	1							
Republic of Guinea	Diecke	45	DIE	2							
South Africa	Cape province	44	CAP			1					
Gabon	Unknown	NA	None	15	18	8					
Republic of Congo	Unknown	NA	None	2		2					
Unknown	Unknown	NA	None	1		1					
	Total			**131**	**475**	**370**	**39**	**5**	**16**	**71**	**39**

5.3.2. Analysis of mitochondrial control region sequences

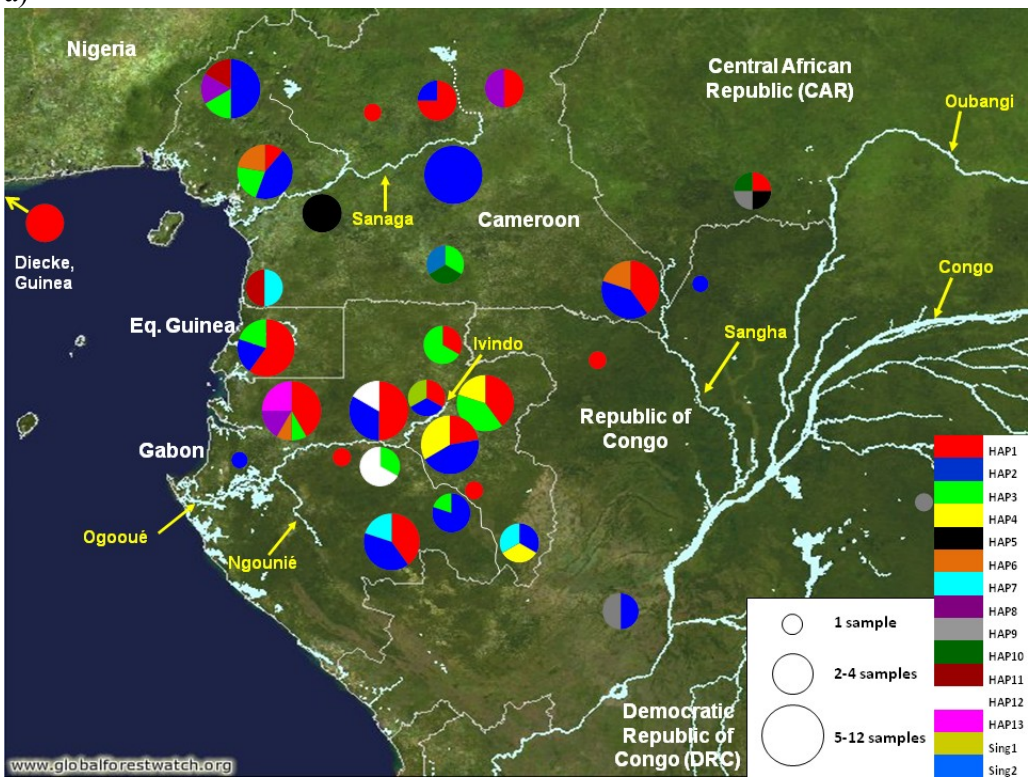
5.3.2.1. Identification of mitochondrial haplogroups

Phylogenetic analysis of *C. dorsalis* sequences revealed 12 reciprocally monophyletic clades with 58% or higher bootstrap support and three sequences of unresolved phylogenetic relationship (see Supplementary Figure 1). Haplogroups 1 and 2 are widespread across the entire sampling area. Haplogroup 3 is widespread across western central Africa (Figure 2a). Two haplogroups are restricted to specific regions: haplogroup 4 is limited to the eastern side of the Ivindo and Ogooué rivers and haplogroup 9 is only found to the eastern side of central Africa (Sites 39, 40, and 42) (Figure 2a). There are also four site-specific haplotypes or haplogroups (Figure 2a): a singleton restricted to central Cameroon (Site 15), a singleton limited to the west of the river Ivindo, a haplogroup (13) that is unique to the upland refugium of the Monts de Cristal (Site 17), and another haplogroup (12) restricted to the North and South banks of the Ogooué river in Ivindo (Sites 25, 26).

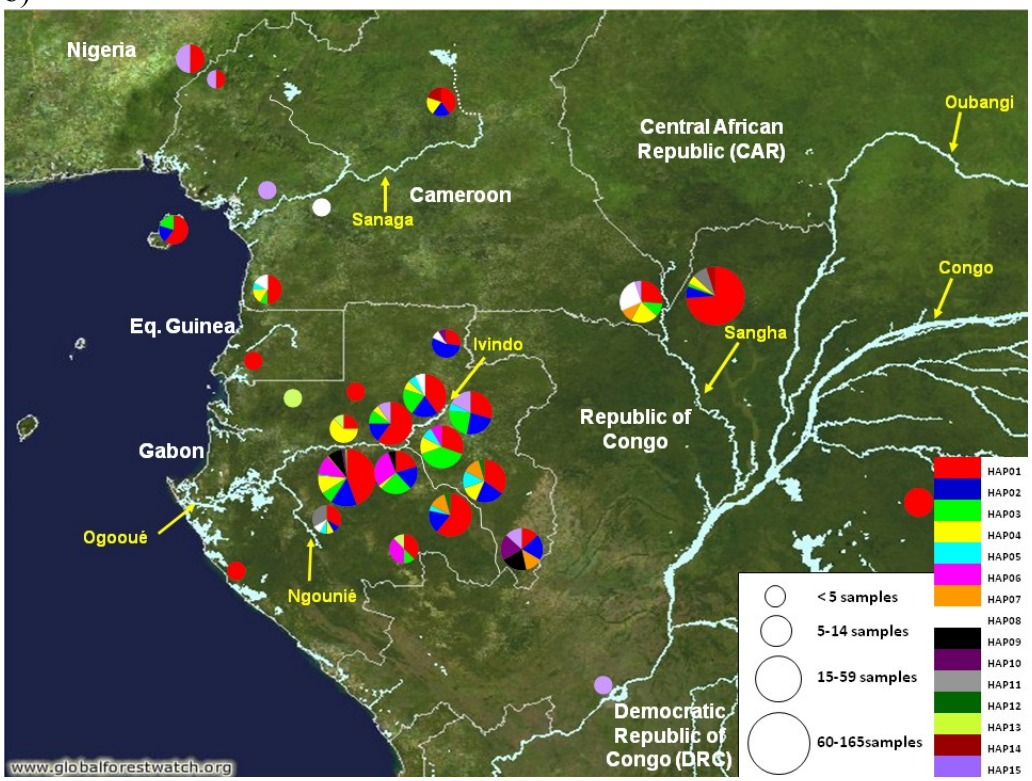
For *C. callipygus*, phylogenetic analysis revealed 14 reciprocally monophyletic clades with 50% or higher bootstrap support and a set of single sequences of unresolved phylogenetic relationship (see Supplementary Figure 2). Most haplogroups are widespread among all sampling sites with the exception of haplogroup 6 which is only found on the south bank of the Ogooué River (Sites 24, 26, 33, and 35) and in Langoué (Site 29) to the north of that same river. Sequences within haplogroup 9 are also only found to the south of Ogooué (Sites 24 and 26) and north of the source of the river (Site 32).

In the case of *P. monticola*, the analysis revealed 19 reciprocally monophyletic clades with 50% or higher bootstrap support and two sets of single sequences of unresolved phylogenetic relationship (see Supplementary Figure 3). The geographic distribution of these haplotypes shows that several are site or region-specific. Haplogroup 10 is solely found in Nigeria (Site 2). Haplogroup 11 is restricted to Cameroon and Equatorial Guinea. Several haplogroups are restricted to Gabon: haplogroup 15 (Sites 17, 24 and 29), haplogroup 17 (Sites 17, 26, 30, and 33), whereas haplogroup 5 and 14 are mostly found south of the Ogooué river, but are also found on the northern bank at OLO_North (Site 30) and HAO (Site 32) close to the source of the Ogooué river. The geographic distribution of a few haplogroups appear to be structured by major central African rivers: haplogroup 2 is only found south of the Sanaga River, haplogroup 12 is found only East of the Ogooué and Ivindo rivers (Sites 29, 32, 38), haplogroup 18 is only found south of the Ogooué (Sites 35 and 33). In addition, haplogroup 1 is mainly found in Cameroon (Sites 6, 14, 3, 5, 7, and 10) with a few exceptions (Sites 1, 2, 20, and 26), haplogroup 16 is only found north of the Ogooué (Sites 17 and 19) and in southeast Cameroon (Site 37), whereas haplogroup 20 is found only in East (Sites 42, 40, 41) and South Africa (Site 44).

a)


Nigeria
Central African Republic (CAR)
Oubangi
Sanaga
Cameroon
Diecke, Guinea
Congo
Eq. Guinea
Sangha
Ivindo
Republic of Congo
Gabon
Ogooué
Ngounié
Democratic Republic of Congo (DRC)
www.globalforestwatch.org
1 sample
2-4 samples
5-12 samples
HAP1
HAP2
HAP3
HAP4
HAP5
HAP6
HAP7
HAP8
HAP9
HAP10
HAP11
HAP12
HAP13
Sing1
Sing2

b)


Nigeria
Central African Republic (CAR)
Oubangi
Cameroon
Sanaga
Congo
Eq. Guinea
Sangha
Ivindo
Republic of Congo
Gabon
Ogooué
Ngounié
Democratic Republic of Congo (DRC)
www.globalforestwatch.org
< 5 samples
5-14 samples
15-59 samples
60-165samples
HAP01
HAP02
HAP03
HAP04
HAP05
HAP06
HAP07
HAP08
HAP09
HAP10
HAP11
HAP12
HAP13
HAP14
HAP15

c)

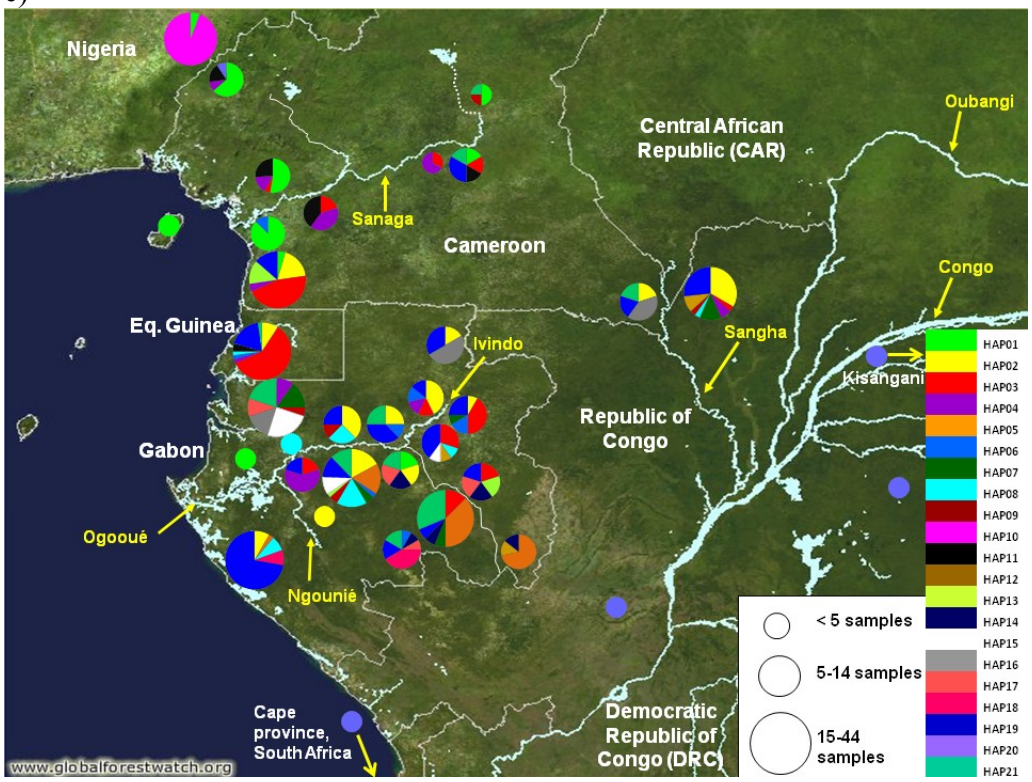
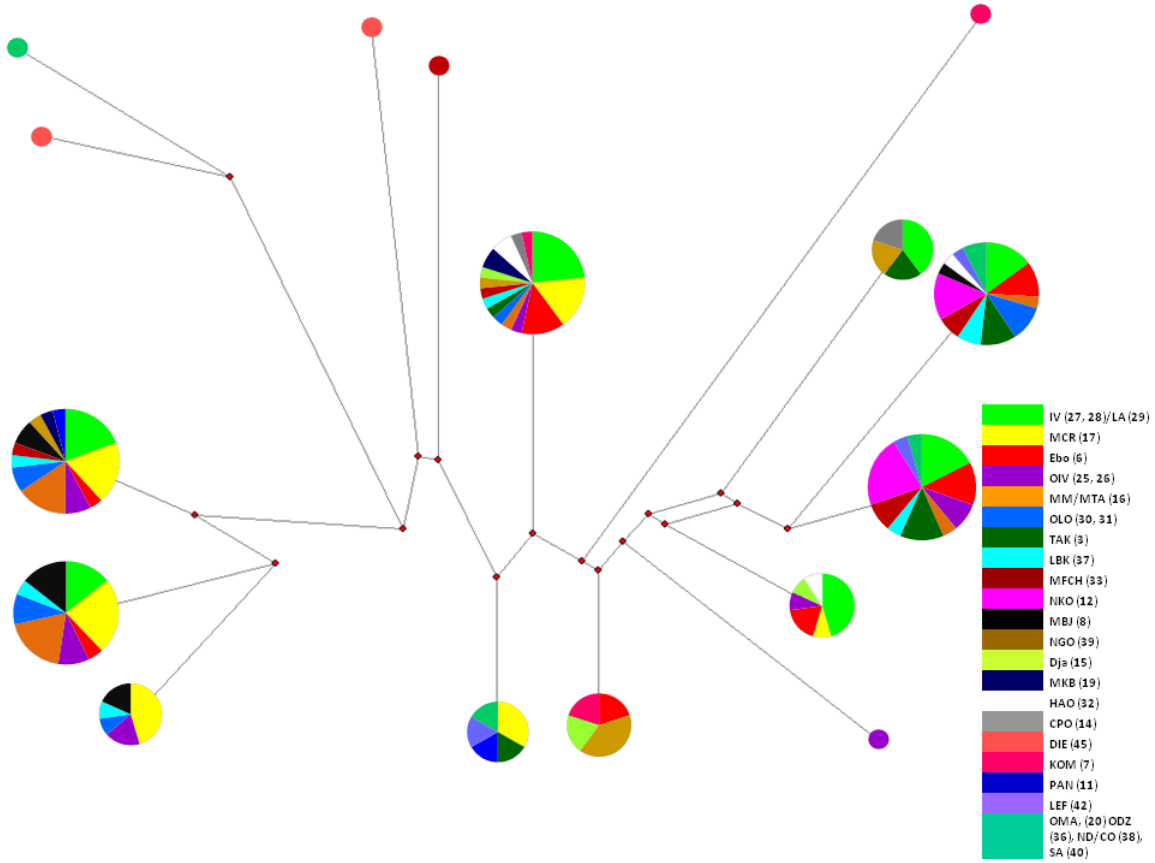


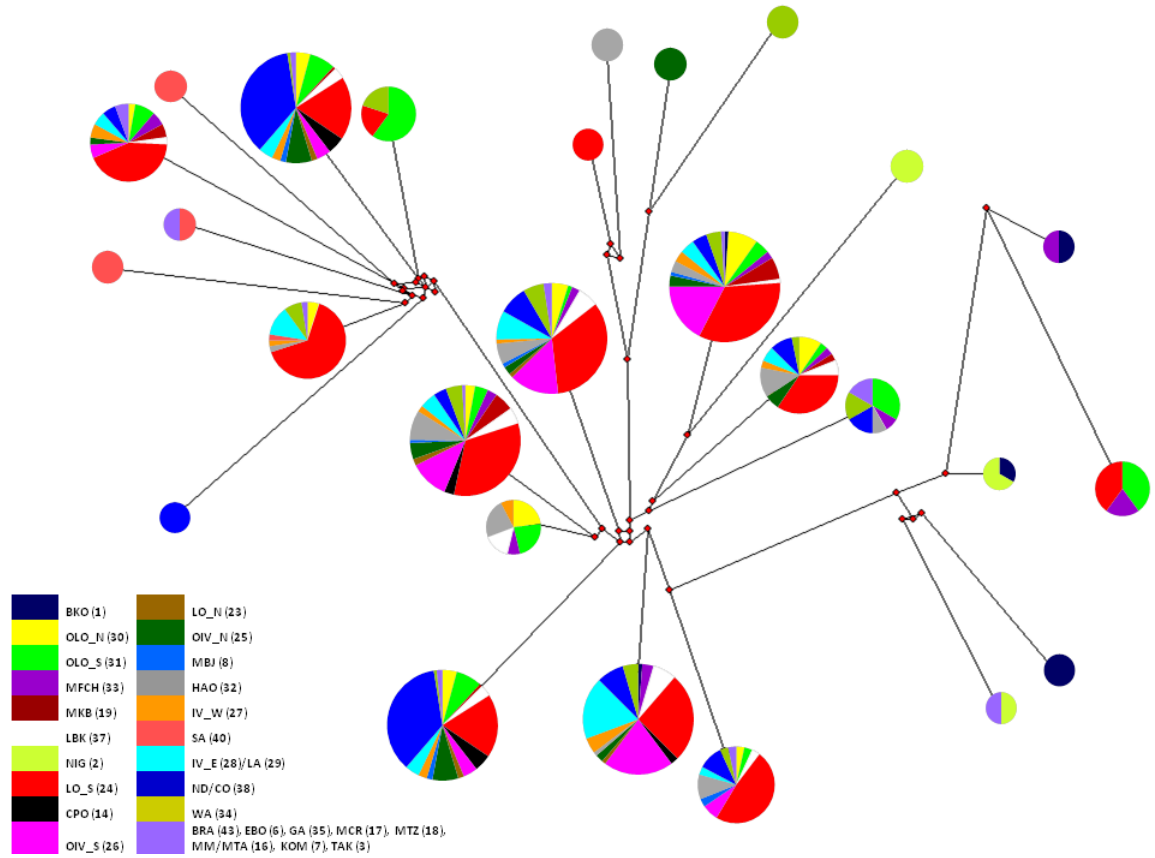
Figure 2. Geographic distribution of (a) *C. dorsalis,* (b) *C. callipygus*, and (c) *P. monticola* haplogroups as defined by phylogenetic analysis of 598 bp, 595 bp, and 757 bp length alignments of the mitochondrial control region, respectively. Map modified from www.globalforestwatch.org.

Within *C. dorsalis*, the Minimum Spanning Network shows that several different haplogroups are widespread among all sampling sites except for one location south of the Sanaga River which is dominated by haplogroup 2 (Figure 3a). Similarly, *C. callipygus* does not reveal any obvious historical pattern of genetic structure (Figure 3b). For *P. monticola*, most of the Nigerian samples form a distinct cluster although a few sequences from this site are also found in the undefined main cluster that is more widespread throughout the region (Figure 3c). Two samples (HAO21 and MM52) belonging to clade B have a 77 bp indel (Ntie *et al.*, 2010a) and are highly differentiated from other haplogroups. GenBank singleton samples from Bamenda (4), East central Africa (41, 42) and South Africa (44) all form a cluster in the network between the Nigerian cluster (2) and the undifferentiated cluster covering much of Central Africa (Figure 3c).

a)


IV (27, 28)/LA (29)
MCR (17)
Ebo (6)
OIV (25, 26)
MM/MTA (16)
OLO (30, 31)
TAK (3)
LBK (37)
MFCH (33)
NKO (12)
MBJ (8)
NGO (39)
Dja (15)
MKB (19)
HAO (32)
CPO (14)
DIE (45)
KOM (7)
PAN (11)
LEF (42)
OMA, (20) ODZ (36), ND/CO (38), SA (40)

b)


BKO (1)
OLO_N (30)
OLO_S (31)
MFCH (33)
MKB (19)
LBK (37)
NIG (2)
LO_S (24)
CPO (14)
OIV_S (26)
LO_N (23)
OIV_N (25)
MBJ (8)
HAO (32)
IV_W (27)
SA (40)
IV_E (28)/LA (29)
ND/CO (38)
WA (34)
BRA (43), EBO (6), GA (35), MCR (17), MTZ (18), MM/MTA (16), KOM (7), TAK (3)

c)

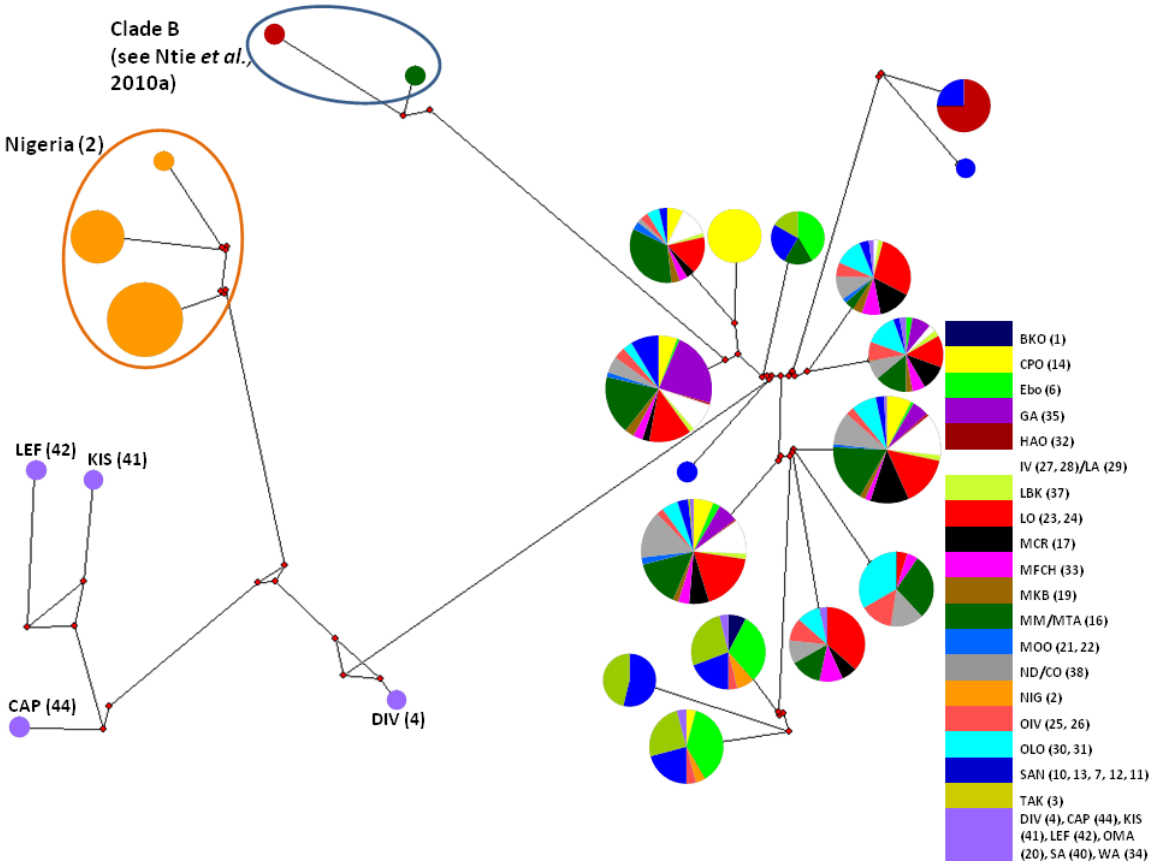


Figure 3. Minimum Spanning Network (MSN) of (a) *C. dorsalis,* (b) *C. callipygus* and (c) *P. monticola* based on 16, 27, and 25 collapsed haplogroups, respectively. Each color represents a different collection site (see legend). The length of the branches is proportional to the number of mutational steps between haplogroups. The size of the pie chart is proportional to the number of samples at a given site. For *C. dorsalis,* 1-4= small; 5-14= medium; 15-30= large. For *C. callipygus,* 1-4= small; 5-19= medium-small; 20-49= medium-large; 50-120= large. For *P. monticola,* 1-3=small; 4-19= medium-small; 20-99= medium-large; 100-200= large. Numbers in parentheses refer to collection sites illustrated in the map of the study area (Figure 1).

5.3.2.2. Spatial and non spatial Analyses of Molecular Variance (SAMOVA and AMOVA)

The *C. dorsalis* SAMOVA analysis shows that the highest F_{CT} value (18.50%) was found when samples were grouped into two clusters (P-value = 0.07) comprising NKO (Site 12), which is midway south of the Sanaga River, and all other remaining sites where this species was sampled (Supplementary Table 2). Aside from this clustering, AMOVA analyses designed to test the possible role of Pleistocene refugia, the Sanaga and Ogooué Rivers showed no significant group structure (Table 2).

Table 2. AMOVA summary statistics. Among group component of the total variance (F_{CT}) for various *C. dorsalis* groupings.

Model	# of groups	F_{CT}	P-value
Monte Mitra/Alen Pleistocene upland refugium: Site (16), Rest	2	4.45	0.229
Monts de Cristal Pleistocene upland refugium: Site (17), Rest	2	-1.53	0.538
Massif du Chaillu Pleistocene upland refugium: Site (33), Rest	2	-9	1
Takamanda Pleistocene upland refugium: Site (3), Rest	2	0.26	0.309
Sangha Pleistocene fluvial refugium: Sites (36, 37, 38, 39), Rest	2	-3.52	0.872
Sanaga Pleistocene fluvial refugium: Sites (5, 7, 8, 9, 11, 12), Rest	2	-1.11	0.58
Ogooué Pleistocene fluvial refugium: Sites (20, 24, 25, 26, 29, 30, 31, 32), Rest	2	-0.95	0.509
All seven Pleistocene refugia: Site (16), Site (17), Site (33), Site (3), Sites (36, 37, 38, 39), Sites (5, 7, 8, 9, 11, 12), Sites (20, 24, 25, 26, 29, 30, 31, 32)	7	-5.83	0.847
Ogooué north: Sites (20, 25, 29, 30, 32), Rest	2	-0.62	0.489
Ogooué south: Sites (24, 26, 31), Rest	2	0.77	0.332
Sanaga north: Sites (6, 8, 9), Rest	2	-3.05	0.865
Sanaga south: Sites (7, 11, 12), Rest	2	2.07	0.196
Ogooué River barrier: Ogooué north (Sites 20, 25, 29, 30, 32), Ogooué south (Sites 24, 26, 31)	2	4.72	0.248
Sanaga River barrier: Sanaga north (Sites 6, 8, 9), Sanaga south (Sites 7, 11, 12)	2	-26.64	0.664
Sanaga and Ogooué River barriers: Sanaga north (Sites 3, 6, 8, 9), Ogooué south (Sites 24, 26, 31, 33), Sanaga south and Ogooué North (Rest)	3	-0.56	0.518

Table 3. AMOVA summary statistics. Among group component of the total variance (F_{CT}) for various *C. callipygus* groupings.

Model	# of groups	F_{CT}	P-value
Massif du Chaillu Pleistocene upland refugium: Site (33), Rest	2	-3.49	0.659
Highlands of Nigeria and Cameroon Pleistocene upland refugium: Site (2, 3), Rest	2	22.06	0.049
Campo Pleistocene lowland refugium: Site (14), Rest	2	-2.52	0.511
Sangha Pleistocene fluvial refugium: Sites (37, 38), Rest	2	4.85	0.113
Sanaga Pleistocene fluvial refugium: Site (6, 7, 8), Rest	2	-6.53	1
Ogooué Pleistocene fluvial refugium: Sites (23, 24, 25, 26, 29, 30, 31, 32), Rest	2	-0.39	0.372
All six Pleistocene refugia: Site (33), Sites (2, 3), Site (14), Sites (37, 38), Sites (6, 7, 8), Sites (23, 24, 25, 26, 29, 30, 31, 32)	6	5.38	0.069
Ogooué north: Sites (23, 25, 29, 30, 32), Rest	2	-0.3	0.401
Ogooué south: Sites (24, 26, 31), Rest	2	-2.46	0.87
Sanaga north: Sites (6, 8), Rest	2	-6.1	1
Ogooué River barrier: Ogooué north (Sites 23, 25, 29, 30, 32), Ogooué south (Sites 24, 26, 31)	2	-1.49	0.703
Sanaga and Ogooué River barriers: Sanaga north (Sites 3, 6, 8, 9), Ogooué south (Sites 24, 26, 31, 33), Sanaga south and Ogooué North (Rest)	3	-0.12	0.379

Table 4. AMOVA summary statistics. Among group component of the total variance (F_{CT}) for various *P. monticola* groupings (NIG excluded).

Model	**# of groups**	**F_{CT}**	**P-value**
Highlands of Nigeria and Cameroon Pleistocene upland refugium: Site (3), Rest	2	34.45	0.082
Monte Mitra/Monte Alen Pleistocene upland refugium: Site (16), Rest	2	-8.66	0.463
Monts de cristal Pleistocene upland refugium: Site (17), Rest	2	-11.87	0.708
Massif du Chaillu Pleistocene upland refugium: Site (33), Rest	2	-9.61	0.498
Campo Pleistocene lowland refugium: Site (14), Rest	2	-11.86	0.746
Sangha Pleistocene fluvial refugium: Sites (37, 38, 39), Rest	2	-0.92	0.354
Sanaga Pleistocene fluvial refugium: Sites (5, 6, 7, 10, 11, 12, 13), Rest	2	19.26	0.007
Ogooué Pleistocene fluvial refugium: Sites (20, 21, 22, 23, 24, 25, 26, 29, 30, 31, 32), Rest	2	-0.51	0.411
All eight Pleistocene refugia: Site (3), Site (16), Site (17), Site (33), Site (14), Sites (37, 38, 39), Sites (5, 6, 7, 10, 11, 12, 13), Sites (20, 21, 22, 23, 24, 25, 26, 29, 30, 31, 32)	8	11	0.056
Ogooué north: Sites (20, 21, 23, 25, 29, 30, 32)	2	-3.1	0.715
Ogooué south: Sites (22, 24, 26, 31)	2	-4.67	0.836
Sanaga south: Sites (5, 7, 10, 11, 12, 13)	2	9.29	0.041
Ogooué River barrier: Ogooué north (Sites 20, 21, 23, 25, 29, 30, 32), Ogooué south (Sites 22, 24, 26, 31)	2	-4.56	0.946
Sanaga River barrier: Sanaga north (Site 6), Sanaga south (Sites 5, 7, 10, 11, 12, 13)	2	-23.44	1
Sanaga and Ogooué River barriers: Sanaga north (Sites 2, 3), Ogooué south (Sites 22, 24, 26, 31, 33, 35), Sanaga south and Ogooué North (Rest)	3	12.64	0.019

SAMOVA analysis of *C. callipygus* also shows that the highest F_{CT} value (19.52%) was found at K=2 (P-value = 0.04), with one group composed of samples from NIG (Site 2) and TAK (Site 3) while the other is made of the remaining sites (Supplementary Table 3). Evidence for the same pattern of population division was also found in *P. monticola* in that NIG (Site 2) forms a very distinct genetic cluster on its own (see above and below). The AMOVA analysis of *C. callipygus* also shows the putative refugium located in the highlands of Nigeria and Cameroon (Sites 2 and 3) is genetically differentiated from the remaining sites (F_{CT} = 22.06%, P-value = 0.049) (Table 3). When the possible effect of other Pleistocene refugia on genetic structure was assessed, there did not appear to be any significant among group F_{CT} component (Table 3). Similarly, neither Sanaga nor Ogooué rivers had any significant effect on the among group variance component F_{CT} (Table 3).

An initial SAMOVA analysis of *P. monticola* shows that the highest F_{CT} value (72.32%, P = 0.039) was found when sites were assembled with one group made up of samples from NIG (Site 2) and the other was made of all other sites. When NIG samples were removed from the dataset, SAMOVA analysis shows again that the highest F_{CT} value (40.57%; P = 0.03) was found at K = 2 (Supplementary Table 4). This grouping now separates Edea (Site 6) on the southern bank of the Sanaga river mouth from the remaining sampling locations. When sites are clustered according to the Sanaga basin a significant among group variance component was also observed (F_{CT} = 19.26%, P = 0.007) (see Table 4). Aside from these results, AMOVA analyses did not reveal any significant group structure, either according to hypothesized refugia or across river banks of the Ogooué and Sanaga (Table 4).

5.3.2.3. Genetic diversity and demographic history

For *C. dorsalis*, several sites have the maximum haplotype diversity possible (Hd =1). These were IV_East (Site 28), LBK (Site 37), MBJ (Site 8), NGO (Site 39), TAK (Site 3), and several sites within the Sangha River basin (Sites 36, 37, 38, and 39). LBK (Site 37) shows the highest nucleotide diversity (π = 0.056) (Table 5). Populations in the Sangha basin (37, 38, 39, 36) have the highest polymorphism per nucleotide site (θ = 0.058) (Table 5). In general tests of departure from neutrality were not significant with the exception of Fu's Fs statistic which suggests a population expansion within haplogroup 2, as illustrated by a negative value and a highly significant p-value (Table 5 and Supplementary Figure 4a). Sequences within this haplogroup are widespread throughout the whole study area with their highest frequency midway south of the Sanaga River in NKO (Site 12) (see Figure 2a and Supplementary Table 5). However, because of the low sample size in NKO (Site 12) (n=5), it is not possible to draw any significant conclusion from the latter result.

For *C. callipygus*, the highest haplotype diversity (Hd =1) was observed in BKO (Site 1) and within the Congo basin in SA (Site 40) (Table 6). BKO (Site 1) is the site with the highest nucleotide diversity (π = 0.079) and polymorphism per nucleotide site (θ = 0.073) (Table 6). The LO_South (Site 24) also has a significant Fu + Li's D* indicative of background selection as a possible cause. Analysis of demographic history within major haplogroups shows a significant Fs value for haplogroup 1 (Table 6 and Supplementary Figure 4b), suggesting a population expansion and/or genetic hitchhiking. In fact, haplogroup 1 is widespread throughout the whole study area with its highest frequency in Northern Congo ND/CO (Site 38) (see Figure 2b and Supplementary Table 5).

In the case of *P. monticola*, the highest haplotype diversity (Hd =1) was found in LBK in Southeastern Cameroon (Site 37), and both IV_East (Site 28) and IV_West (Site 27) in

central Gabon (Table 7). Nine individual haplogroups also exhibit maximal haplotype diversity (Hd =1). These are: haplogroup 7 (Sites 17, 28, 38, 24, and 31), haplogroup 9 (Sites 24, 17, 38, and 10), haplogroup 10 (Site 2), haplogroup 12 (Sites 38, 32, and 29), haplogroup 13 (Sites 24, 14), haplogroup 14 (Sites 26, 32, and 33), haplogroup 15 (Sites 17, 29, and 24), haplogroup 17 (Sites 17, 33, and 30), haplogroup 18 (Sites 35, 33). The highest average nucleotide diversity ($\pi = 0.076$) was found at Ebo (Site 6) and highest polymorphism per nucleotide site ($\theta = 0.08$) in CPO (Site 14) (Table 7). Analysis of demographic history shows significant Fs and/or R^2 p-value for haplogroups 1, 2, 3, and 9 (Table 7 and Supplementary Figure 4c), suggesting a history of past population expansion and/or genetic hitchhiking. Haplogroups 1, 2, and 3 have their highest frequency in Edea (5), ND/CO (37), and MM/MTA (16) respectively (See Figure 2c and Supplementary Table 5).

Table 5. *C. dorsalis* summary statistics of genetic diversity indices, neutrality tests, and test of population growth.

Site names (site #)	n	Hd	π	θ	Tajima's D	Fu and Li's D*	Fu and Li's F*	Fu's Fs	R^2
Ebo (6)	9	0.944	0.03883	0.04021	P > 0.10	P > 0.10	P > 0.10	NA	NA
IV_East (28)	5	1	0.04802	0.0446	P > 0.10	P > 0.10	P > 0.10	NA	NA
LA (29)	9	0.972	0.03647	0.03575	P > 0.10	P > 0.10	P > 0.10	NA	NA
LBK (37)	5	1	0.05585	0.054	P > 0.10	P > 0.10	P > 0.10	NA	NA
MBJ (8)	4	1	0.04243	0.04477	P > 0.10	P > 0.10	P > 0.10	NA	NA
MCR (17)	12	0.788	0.04343	0.03341	P > 0.10	P > 0.10	P > 0.10	NA	0.2082
MFCH (33)	5	0.9	0.04783	0.04666	P > 0.10	P > 0.10	P > 0.10	NA	NA
MM/MTA (16)	6	0.933	0.03846	0.03669	P > 0.10	P > 0.10	P > 0.10	NA	NA
NGO (39)	4	1	0.04487	0.04561	P > 0.10	P > 0.10	P > 0.10	NA	NA
NKO (12)	5	0.9	0.00318	0.00332	P > 0.10	P > 0.10	P > 0.10	NA	NA
OIV_North (25)	6	0.933	0.05006	0.04401	P > 0.10	P > 0.10	P > 0.10	NA	NA
OLO_North (30)	4	0.5	0.02425	0.02645	P > 0.10	P > 0.10	P > 0.10	NA	NA
TAK (3)	6	1	0.03657	0.03742	P > 0.10	P > 0.10	P > 0.10	NA	NA
Ogooue north (20, 25, 29, 30, 32)	20	0.974	0.04447	0.03966	P > 0.10	P > 0.10	P > 0.10	0.4327	0.1493
Ogooue south (24, 26, 31),	9	0.889	0.04182	0.04129	P > 0.10	P > 0.10	P > 0.10	NA	NA
Sanaga north (6, 8, 9)	14	0.956	0.04906	0.03973	P > 0.10	P > 0.10	P > 0.10	NA	0.1869
Sanaga south (7, 11, 12)	9	0.889	0.04286	0.05067	P > 0.10	P > 0.10	P > 0.10	NA	NA
Ogooué basin (20, 24, 25, 26, 29, 30, 31, 32)	29	0.973	0.04541	0.03923	P > 0.10	P > 0.10	P > 0.10	0.1525	NA
Sanaga basin (5, 7, 8, 9, 11, 12)	23	0.968	0.04793	0.04937	P > 0.10	P > 0.10	P > 0.10	0.3624	NA
Sangha basin (37, 38, 39, 36)	11	1	0.05464	0.05777	P > 0.10	P > 0.10	P > 0.10	NA	0.1276
Haplogroup 1 (36, 33, 39, 8, 6, 30, 29, 37, 23, 17, 16, 25, 11, 28, 19)	27	0.952	0.03185	0.04298	P > 0.10	P > 0.10	P > 0.10	-1.9501	NA
Haplogroup 2 (32, 38, 8, 6, 3, 20, 42, 27, 33, 37, 31, 29, 25, 16, 12)	27	0.986	0.02051	0.02091	P > 0.10	P > 0.10	P > 0.10	-7.0579 *	NA
Haplogroup 3 (15, 28, 26, 31, 19, 6, 17, 16, 3)	10	0.933	0.00931	0.00828	P > 0.10	P > 0.10	P > 0.10	NA	0.1642
Haplogroup 4 (28, 29, 32)	5	0.9	0.0081	0.00883	P > 0.10	P > 0.10	P > 0.10	NA	NA

Asterisk indicates significant correlation: *=P < 0.05, **=P < 0.01, ***=P < 0.001.

NA: not applicable because of sample size inferior to 20 for Fu's Fs and sample size less than 10 or more than 20 for R^2 (Ramos-Onsins, Rozas, 2002).

Haplogroups 1 through 4 correspond to major Haplogroups as illustrated in Figure 2a and Supplementary Figure 1.

Table 6. *C. callipygus* summary statistics of genetic diversity indices, neutrality tests, and test of population growth.

Site names	n	Hd	π	θ	Fu + Li's D*	Fu + Li's F*	Tajima's D	Fu's Fs	R^2
BKO (1)	5	1	0.07877	0.07315	P > 0.10	P > 0.10	P > 0.10	NA	NA
CPO (14)	12	0.939	0.05248	0.04669	P > 0.10	P > 0.10	P > 0.10	NA	0.715
HAO (32)	15	0.952	0.06248	0.0664	P > 0.10	P > 0.10	P > 0.10	NA	0.367
IV_East (28)	17	0.985	0.06033	0.05793	P > 0.10	P > 0.10	P > 0.10	NA	0.146
IV_West (27)	12	0.939	0.0712	0.06412	P > 0.10	P > 0.10	P > 0.10	NA	0.686
LA (29)	21	0.957	0.05506	0.04688	P > 0.10	P > 0.10	P > 0.10	1.75	NA
LBK (37)	16	0.983	0.06731	0.06988	P > 0.10	P > 0.10	P > 0.10	NA	0.399
LO_North (23)	8	0.821	0.04378	0.03785	P > 0.10	P > 0.10	P > 0.10	NA	0.846
LO_South (24)	135	0.981	0.07369	0.06219	P < 0.05	P > 0.10	P > 0.10	0.282	NA
MBJ (8)	5	0.9	0.06603	0.06265	P > 0.10	P > 0.10	P > 0.10	NA	NA
MFCH (33)	8	0.929	0.07064	0.06877	P > 0.10	P > 0.10	P > 0.10	NA	NA
MKB (19)	11	0.945	0.04564	0.04836	P > 0.10	P > 0.10	P > 0.10	NA	0.264
ND/CO (38)	60	0.972	0.04773	0.04919	P > 0.10	P > 0.10	P > 0.10	0.47	NA
NIG (2)	6	0.733	0.05419	0.04579	P > 0.10	P > 0.10	P > 0.10	NA	NA
OIV_North (25)	20	0.942	0.06733	0.05822	P > 0.10	P > 0.10	P > 0.10	0.968	0.826
OIV_South (26)	34	0.918	0.05382	0.04991	P > 0.10	P > 0.10	P > 0.10	0.971	NA
OLO_North (30)	23	0.933	0.06008	0.05531	P > 0.10	P > 0.10	P > 0.10	0.987	NA
OLO_South (31)	23	0.953	0.06639	0.06059	P > 0.10	P > 0.10	P > 0.10	0.895	NA
SA (40)	5	1	0.06652	0.06918	P > 0.10	P > 0.10	P > 0.10	NA	NA
WA (34)	11	0.982	0.06606	0.06738	P > 0.10	P > 0.10	P > 0.10	NA	0.325
Ogooué north (23, 25, 29, 30, 32)	87	0.984	0.06032	0.05864	P > 0.10	P > 0.10	P > 0.10	-3.05	NA
Ogooué south (24, 26, 31)	192	0.985	0.06032	0.05463	P > 0.10	P > 0.10	P > 0.10	-9.7646	NA
Sanaga north (6, 8)	6	0.933	0.06239	0.05746	P > 0.10	P > 0.10	P > 0.10	NA	NA
Ogooué basin (23, 24, 25, 26, 29, 30, 31, 32)	279	0.989	0.06057	0.05758	P > 0.10	P > 0.10	P > 0.10	-26.9774**	NA
Sanaga Basin (6, 7, 8)	7	0.952	0.06137	0.05827	P > 0.10	P > 0.10	P > 0.10	NA	NA
Sangha Basin (37, 38)	76	0.977	0.05891	0.05993	P > 0.10	P > 0.10	P > 0.10	-2.736	NA
Haplogroup 1 (2, 1, 3, 31, 24, 33, 38, 40, 37, 18, 19, 25, 34, 8, 28, 30, 29, 32, 35, 16, 27)	206	0.989	0.0493	0.06406	P > 0.10	P > 0.10	P > 0.10	-38.4795***	NA
Haplogroup 2 (1, 28, 24, 38, 32, 8, 19, 25, 26, 30, 34)	61	0.951	0.03313	0.04692	P > 0.10	P > 0.10	P > 0.10	-0.332	NA
Haplogroup 3 (26, 14, 29, 1, 24, 38, 25, 37, 33)	41	0.839	0.02662	0.02093	P > 0.10	P > 0.10	P > 0.10	1.6703	NA
Haplogroup 4 (8, 37, 14, 29, 24, 25, 30, 34, 38)	39	0.846	0.01898	0.0253	P > 0.10	P > 0.10	P > 0.10	1.2943	NA
Haplogroup 5 (28, 24, 34, 29, 30)	11	0.945	0.02589	0.03072	P > 0.10	P > 0.10	P > 0.10	NA	0.1198
Haplogroup 6 (35, 24, 33, 29, 26)	31	0.905	0.02413	0.02581	P > 0.10	P > 0.10	P > 0.10	1.0238	NA
Haplogroup 7 (37, 31, 32, 30)	11	0.873	0.03182	0.03176	P > 0.10	P > 0.10	P > 0.10	NA	0.1651

Table 6 (continued). *C. callipygus* summary statistics of genetic diversity indices, neutrality tests, and test of population growth.

Haplogroup 8 (14, 19, 37, 7)	11	0.945	0.02961	0.03013	P > 0.10	P > 0.10	P > 0.10	NA	0.1374
Haplogroup 9 (26, 24, 32)	16	0.825	0.02312	0.02344	P > 0.10	P > 0.10	P > 0.10	NA	0.1563
Haplogroup 10 (32, 24, 19)	6	0.933	0.03716	0.03801	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 11 (38, 34)	9	0.917	0.03951	0.04661	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 13 (17, 24, 33)	4	0.833	0.01774	0.01935	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 14 (38, 8)	4	0.5	0.00169	0.00184	P > 0.10	P > 0.10	P > 0.10	NA	NA

Asterisk indicates significant correlation: *=P < 0.05, **=P < 0.01, ***=P < 0.001.
NA: not applicable because of sample size inferior to 20 for Fu's Fs and sample size less than 10 or more than 20 for R^2 (Ramos-Onsins, Rozas, 2002).
Haplogroups 1 through 14 correspond to major Haplogroups as illustrated in Figure 2b and Supplementary Figure 2.

Table 7. *P. monticola* summary statistics of genetic diversity indices, neutrality tests, and test of population growth.

Site names	**n**	**Hd**	**π**	**θ**	**Fu + Li's D***	**Fu + Li's F***	**Tajima's D**	**Fu's Fs**	**R^2**
CPO (14)	21	0.986	0.06636	0.07995	P > 0.10	P > 0.10	P > 0.10	-2.3216	NA
Ebo (6)	19	0.942	0.07556	0.0594	P > 0.10	P > 0.10	P > 0.10	NA	0.1727
Edea (5)	8	0.964	0.04517	0.05369	P > 0.10	P > 0.10	P > 0.10	NA	NA
GA (35)	33	0.769	0.03376	0.03096	P > 0.10	P > 0.10	P > 0.10	11.5556	NA
HAO (32)	7	0.857	0.03752	0.04288	P > 0.10	P > 0.10	P > 0.10	NA	NA
IV_East (28)	9	1	0.04045	0.04482	P > 0.10	P > 0.10	P > 0.10	NA	NA
IV_West (27)	7	1	0.05145	0.05136	P > 0.10	P > 0.10	P > 0.10	NA	NA
KOM (7)	5	0.9	0.05313	0.04736	P > 0.10	P > 0.10	P > 0.10	NA	NA
LA (29)	6	0.8	0.04044	0.04064	P > 0.10	P > 0.10	P > 0.10	NA	NA
LBK (37)	4	1	0.04585	0.04678	P > 0.10	P > 0.10	P > 0.10	NA	NA
LO_North (23)	7	0.857	0.03783	0.03728	P > 0.10	P > 0.10	P > 0.10	NA	NA
LO_South (24)	33	0.981	0.04314	0.05187	P > 0.10	P > 0.10	P > 0.10	-1.1426	NA
MCR (17)	16	0.95	0.05685	0.05485	P > 0.10	P > 0.10	P > 0.10	NA	0.1445
MFCH (33)	9	0.972	0.05381	0.05681	P > 0.10	P > 0.10	P > 0.10	NA	NA
MKB (19)	4	0.833	0.03743	0.03807	P > 0.10	P > 0.10	P > 0.10	NA	NA
MM/MTA (16)	35	0.96	0.04687	0.05413	P > 0.10	P > 0.10	P > 0.10	1.2536	NA
MOO_South (22)	5	0.7	0.04153	0.04294	P > 0.10	P > 0.10	P > 0.10	NA	NA
ND/CO (38)	30	0.993	0.05185	0.06322	P > 0.10	P > 0.10	P > 0.10	-5.4148*	NA
NIG (2)	35	0.931	0.073	0.07384	P > 0.10	P > 0.10	P > 0.10	4.6251	NA
NKO (12)	6	0.933	0.0693	0.07333	P > 0.10	P > 0.10	P > 0.10	NA	NA

Table 7 (continued). *P. monticola* summary statistics of genetic diversity indices, neutrality tests, and test of population growth.

OIV_North (25)	4	0.833	0.04193	0.04082	P > 0.10	P > 0.10	P > 0.11	NA	NA
OLO_South (26)	9	0.917	0.03691	0.03909	P > 0.10	P > 0.10	P > 0.10	NA	NA
TAK (3)	10	0.711	0.05777	0.05401	P > 0.10	P > 0.10	P > 0.10	NA	0.1677
GA (35)	33	0.769	0.03376	0.03096	P > 0.10	P > 0.10	P > 0.10	11.5556	NA
NIG (2)	35	0.931	0.073	0.07384	P > 0.10	P > 0.10	P > 0.10	4.6251	NA
Ogooué north (20, 21, 23, 25, 29, 30, 32)	28	0.979	0.04561	0.05827	P > 0.10	P > 0.10	P > 0.10	-1.5913	NA
Ogooué south (22, 24, 26, 31)	50	0.986	0.04611	0.06067	P > 0.10	P > 0.10	P > 0.10	-5.1577	NA
Sanaga south (5, 7, 10, 11, 12, 13)	78	0.99	0.04533	0.06046	P > 0.10	P > 0.10	P > 0.10	-13.3752**	NA
Ogooué basin (20, 21, 22, 23, 24, 25, 26, 29, 30, 31, 32)	26	0.988	0.07345	0.06567	P > 0.10	P > 0.10	P > 0.10	-2.4354	NA
Sanaga basin (5, 6, 7, 10, 11, 12, 13)	45	0.986	0.07024	0.07024	P > 0.10	P > 0.10	P > 0.10	-5.6155*	NA
Sangha Basin (37, 38)	34	0.995	0.05216	0.06537	P > 0.10	P > 0.10	P > 0.10	-7.6718*	NA
Haplogroup 1 (10, 3, 6, 1, 12, 14, 2, 20, 26, 5)	32	0.992	0.03664	0.04949	P > 0.10	P > 0.10	P > 0.10	-11.8919 ***	NA
Haplogroup 2 (38, 37, 24, 16, 34, 19, 14, 28, 25)	32	0.994	0.02417	0.03849	P > 0.10	P > 0.10	P > 0.10	-13.3733 ***	NA
Haplogroup 3 (38, 14, 6, 28, 16, 22, 30, 7, 29)	39	0.997	0.02792	0.04301	P > 0.10	P > 0.10	P > 0.10	-21.7542 ***	NA
Haplogroup 4 (7, 22, 38, 16, 13, 27, 6, 3)	13	0.974	0.02932	0.03348	P > 0.10	P > 0.10	P > 0.10	NA	0.1127
Haplogroup 5 (35, 32, 24, 31)	12	0.985	0.02902	0.03135	P > 0.10	P > 0.10	P > 0.10	NA	0.1075
Haplogroup 6 (25, 5, 27, 24, 28, 33)	8	0.964	0.02545	0.02637	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 7 (17, 28, 38, 24, 31)	9	1	0.02537	0.0277	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 8 (35, 38, 24, 16, 21)	14	0.989	0.03067	0.03309	P > 0.10	P > 0.10	P > 0.10	NA	0.1108
Haplogroup 9 (24, 17, 38, 10)	5	1	0.02246	0.02338	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 10 (2)	17	1	0.05537	0.04784	P > 0.10	P > 0.10	P > 0.10	NA	0.1538
Haplogroup 11 (6, 16, 7, 3)	7	0.952	0.01042	0.01203	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 12 (38, 32, 29)	5	1	0.0143	0.01497	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 13 (24, 14)	5	1	0.02932	0.02632	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 14 (26, 32, 33)	5	1	0.02856	0.02737	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 15 (17, 29, 24)	8	1	0.02243	0.02341	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 16 (17, 37, 19)	7	0.952	0.03059	0.0293	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 17 (17, 33, 30)	4	1	0.02801	0.02663	P > 0.10	P > 0.10	P > 0.10	NA	NA
Haplogroup 18 (35, 33)	6	1	0.02868	0.03158	P > 0.10	P > 0.10	P > 0.10	NA	NA

Asterisk indicates significant correlation: *=P < 0.05, **=P < 0.01, ***=P < 0.001.

NA: not applicable because of sample size inferior to 20 for Fu's Fs and sample size inferior to 10 or more than 20 for R^2 (Ramos-Onsins, Rozas, 2002).

Haplogroups 1 through 21 correspond to major haplogroups as illustrated in Figure 2c and Supplementary Figure 3 .

5.3.3. Analysis of nuclear microsatellites

5.3.3.1. Expected heterozygosity and test of Linkage Equilibrium (LE) and Hardy-Weinberg Equilibrium (HWE)

In *C. dorsalis*, samples from Ebo (6) and OLO_South (Site 24) have the highest expected heterozygosity (He = 0.77) (Supplementary Table 6). After Bonferroni correction, there is no locus at any site in linkage disequilibrium and all sites are in HWE (Supplementary Table 6). For *C. callipygus*, results show that IV_West (Site 27) had the highest mean observed heterozygosity (He = 0.74) (Supplementary Table 7). In addition five sites had one pair of loci significantly in linkage disequilibrium after Bonferroni correction: IV_West (Site 27) (BM2113, INRA40), LO_North (Site 23) (SR12, BM143), OIV_North (Site 25) (SR12, INRA05), OLO_North (Site 30) (BM121, INRA05), and OLO_South (31) (INRA40, SR12) (Supplementary Table 8). The loci SR12 and BM864 are also significantly in Hardy-Weinberg at one site (Site 24) after Bonferroni correction (Supplementary Table 7). Lastly, the highest expected heterozygosity (He = 0.87) for *P. monticola* was observed at Ebo (Site 6) (Supplementary Table 9). This site also has one pair of loci (BM2113 and INRA40) significantly in linkage disequilibrium after Bonferroni correction (Supplementary Table 10). LO_South (Site 24) has two pairs of loci (BM2113 and INRA40; BM2113 and BM1225) significantly in linkage disequilibrium after Bonferroni correction (Supplementary Table 10). Finally, CPO (Site 14) has one locus (MM12) in Hardy-Weinberg disequilibrium, LO_South (Site 24) has one (BM121), and NGO (Site 39) has three (BM121, BM848, and SR12) after Bonferroni correction (Supplementary Table 9).

5.3.3.2. Probability of Identity (P_{ID}) and relatedness (r) among samples

Gimlet analyses show that the minimum number of typed microsatellite loci needed to identify first-order relatives is five for *C. dorsalis*, six for *C. callipygus,* and four for *P. monticola*. KINGROUP found 11 out of 65, 77 out of 310, and 82 out 332 samples that were first-order relatives for *C. dorsalis*, *C. callipygus,* and *P. monticola* respectively.

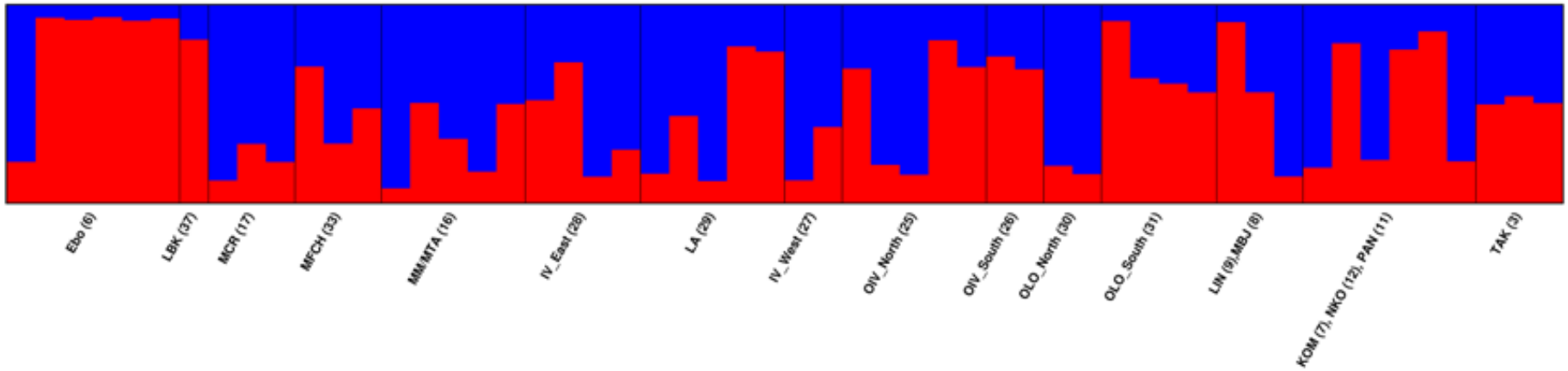
5.3.3.3. Population genetic structure with the program STRUCTURE

Results of the STRUCTURE analyses for *C. dorsalis* indicate that with first-order relatives included in the dataset, the best estimate of K is 3 (Supplementary Figure 5a) and 2 without first-order relatives (Figure 4a). Regardless, in both cases populations seem to be grouped into one cluster encompassing sites in Cameroon and a second made up sites in Gabon and Equatorial Guinea. However, even this grouping is not coherent because there is no clear genetic differentiation between the two regions and several Gabonese sites (26 and 31) are similar to sites in Cameroon (Figure 4a).

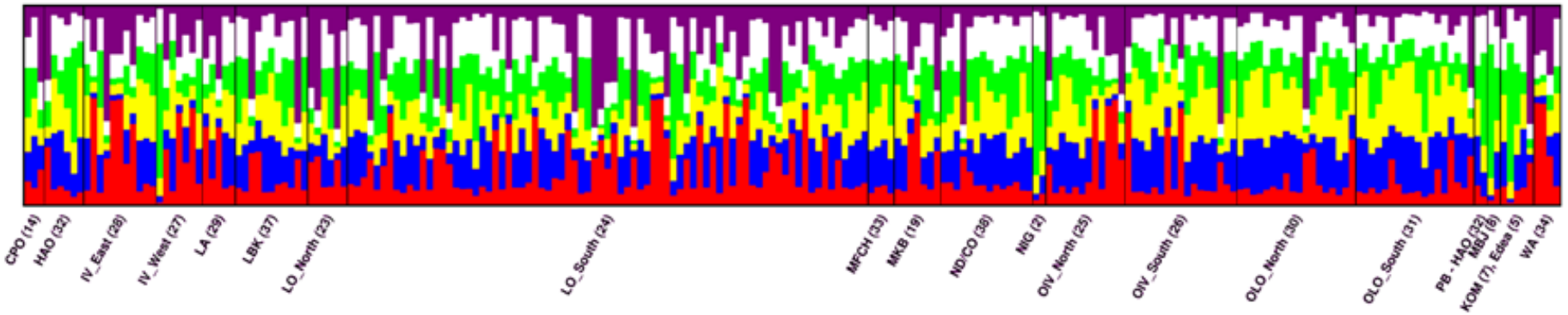
In *C. callipygus*, results of STRUCTURE analyses with and without first-order relatives also show different best K values. When first-order relatives are included, the best K is three (Supplementary Figure 5b). When first-order relatives were excluded, K values varied between 3, 4 and 6 (see Figure 4b for K= 6). However, there was no obvious geographic differentiation despite the observation that the raw mean estimated likelihood probability of data at K=1 is not the highest, indicating that a model of population subdivision is more plausible.

P. monticola analyses, with and without first-order relatives, indicate that the best K is 2, with NGO (39) being differentiated from all other populations (see Figure 4c and Supplementary Figure 5c). When STRUCTURE was run on the same dataset without the NGO (39) samples, two different estimates of K were obtained depending on whether first-order relatives were included (K=3) or not (K=2). When first-order relative were included (K=3), GA (35) appears differentiated (Supplementary Figure 5d) but this is not the case when first-order relatives are excluded (K=2) (Figure 4d). In fact, when K was set to 2, all populations seem to be equally admixed. The latter result is supported by a highest raw mean estimated likelihood probability of data at K=1, suggesting no genetic differentiation among assessed populations once NGO (39) has been removed. It then appears that having first-order relatives in the dataset influences the outcome of the STRUCTURE analyses. These findings were then corroborated with the program GENELAND which takes into account both geographic location and genetic information of each sample.

a)


Ebo (6)
LBK (37)
MCR (17)
MFCH (33)
MM;MTA (16)
IV_East (28)
LA (29)
IV_West (27)
OIV_North (25)
OIV_South (26)
OLO_North (30)
OLO_South (31)
LIN (9);MBJ (8)
KOM (7); NKO (12); PAN (11)
TAK (3)

b)


CPD (14)
HAO (32)
IV_East (28)
IV_West (27)
LA (29)
LBK (37)
LO_North (23)
LO_South (24)
MFCH (33)
MKB (19)
NDCO (38)
NIG (2)
OIV_North (25)
OIV_South (26)
OLO_North (30)
OLO_South (31)
PB - HAO (32)
MBJ (8)
KOM (7); Edea (5)
WA (34)

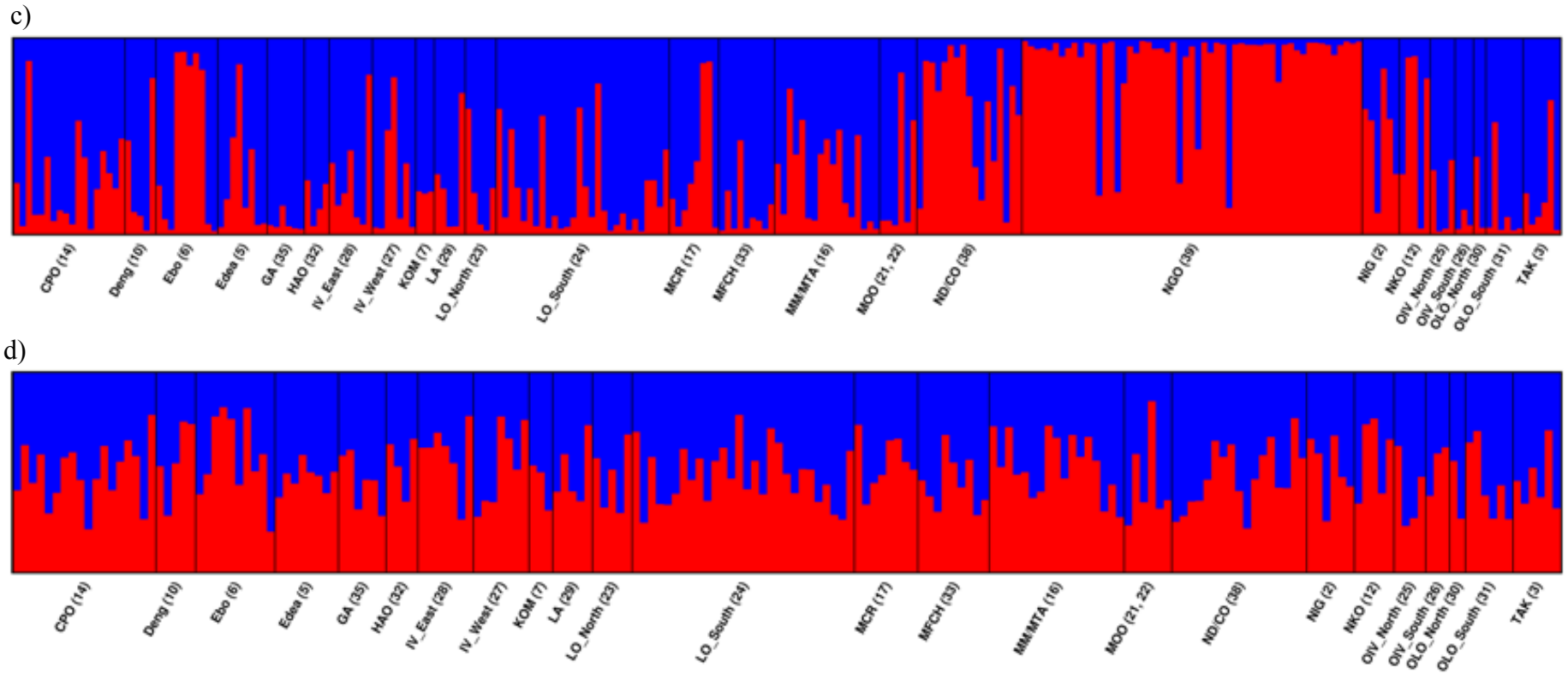


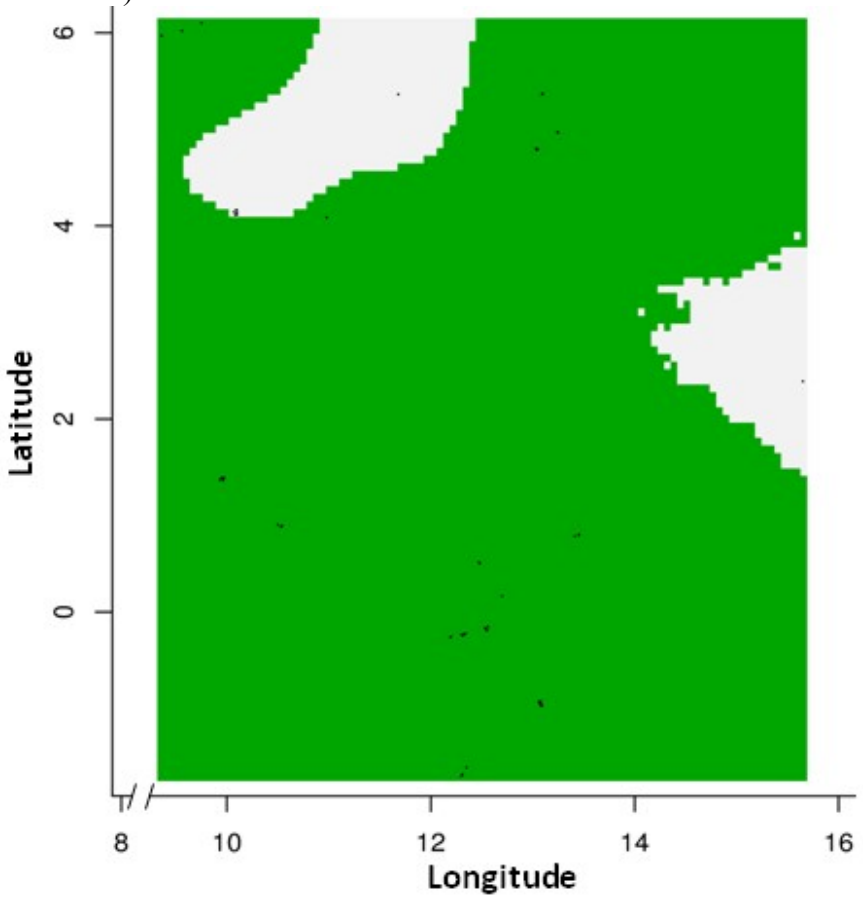
Figure 4. Graphical representation of the STRUCTURE output of (a) *C. dorsalis* (b) *C. callipygus* (c) *P. monticola* without first order relatives respectively with the best K estimated to be equal to 2, 6, 2 respectively. Lastly (d) *P. monticola* is illustrated with NGO excluded. Color graphics represent the proportional membership of each individual genotype to a given cluster. Numbers in parentheses refer to collection sites illustrated in the map of the study area (Figure 1).

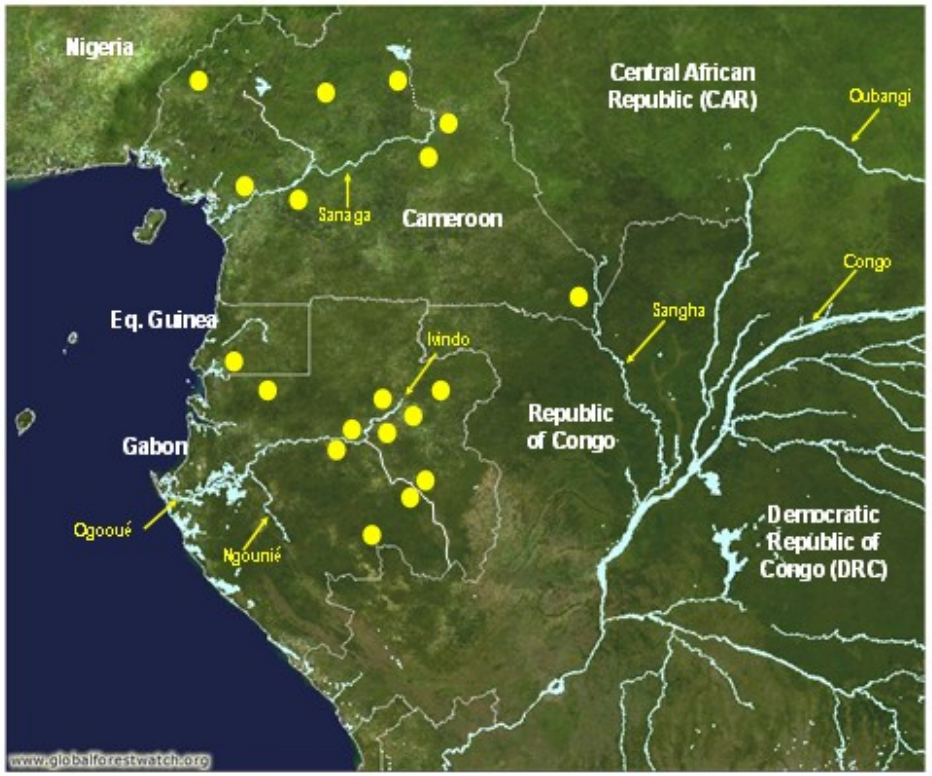
5.3.3.4. Population genetic structure with the program GENELAND

For *C. dorsalis*, GENELAND analyses indicated that the best value of K was two, regardless of whether first order relatives were included or not. The analysis reveals one cluster composed of Ebo (Site 6) and a second made up of all other sites. However, one singleton from LBK (Site 37) in southeastern Cameroon is assigned to both clusters (Supplementary Figure 6a). In the analysis without first-order relatives, one cluster is made up of Ebo (Site 6), LIN (Site 9) and LBK (Site 37) in Cameroon while the other cluster comprises the remaining populations (Figure 5a). These results suggest a possible barrier effect of the Sanaga River on *C. dorsalis*, since Ebo (Site 6) and LIN (Site 9) are located on its northern bank. However, the low sample size and the location of LBK (Site 37) southeast of the Sanaga River make it difficult to draw any significant conclusions from the present analysis.

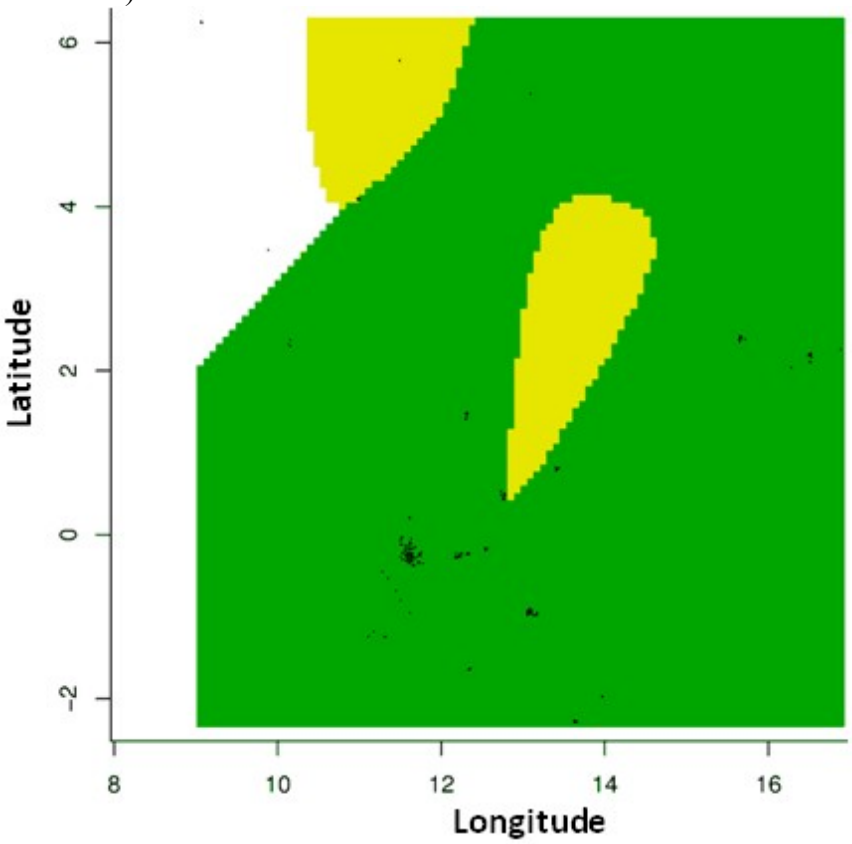
Analyses of *C. callipygus* with and without first-order relatives shows that the best number of K clusters is 3. On one hand, when first-order relatives are included, clusters one and three overlap almost entirely (Supplementary Figures 6b) and membership probabilities among individuals for all three clusters vary between 0.29 and 0.38. Additionally, there is one sample (IP16) from IV_West (Site 27) that belongs equally to cluster one and three (Supplementary Figures 6b). These results suggest that all individuals are equally likely to belong to any of the three clusters and are equally admixed, giving little support to any historical population structure. On the other hand, the analysis without first-order relatives defines three non-overlapping clusters (Figure 5b): cluster two comprises two samples (San005 and San044) from opposite banks of the Sanaga River (Sites 9 and 7 respectively); cluster 3 comprises two NIG (Site 2) samples (Nig22 and Nig67) and one sample from Edea to the South of the Sanaga (Site 5) (San066); and cluster 1 comprises the rest of the samples (Figure 5b). Furthermore, there are two samples (IP16 and IP30) from IV_West (Site 27) that belong to both cluster one (membership probabilities of 0.52 and 0.27 respectively) and two (membership probabilities 0.27 and 0.3 respectively) (Figure 5b).

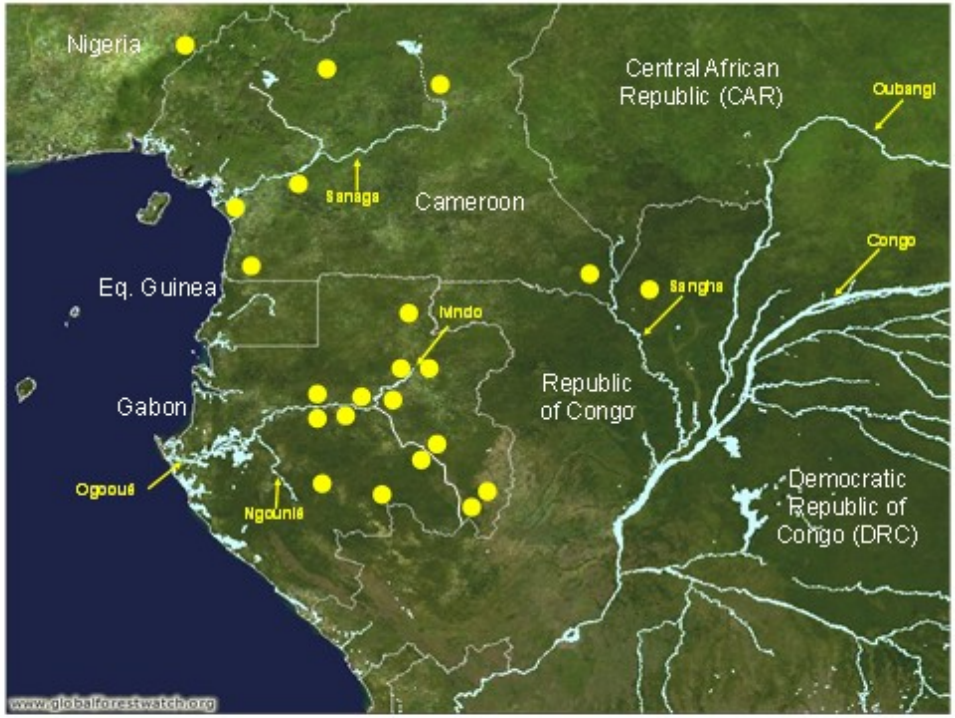
a)


Latitude
6
4
2
0
8
10
12
14
16
Longitude

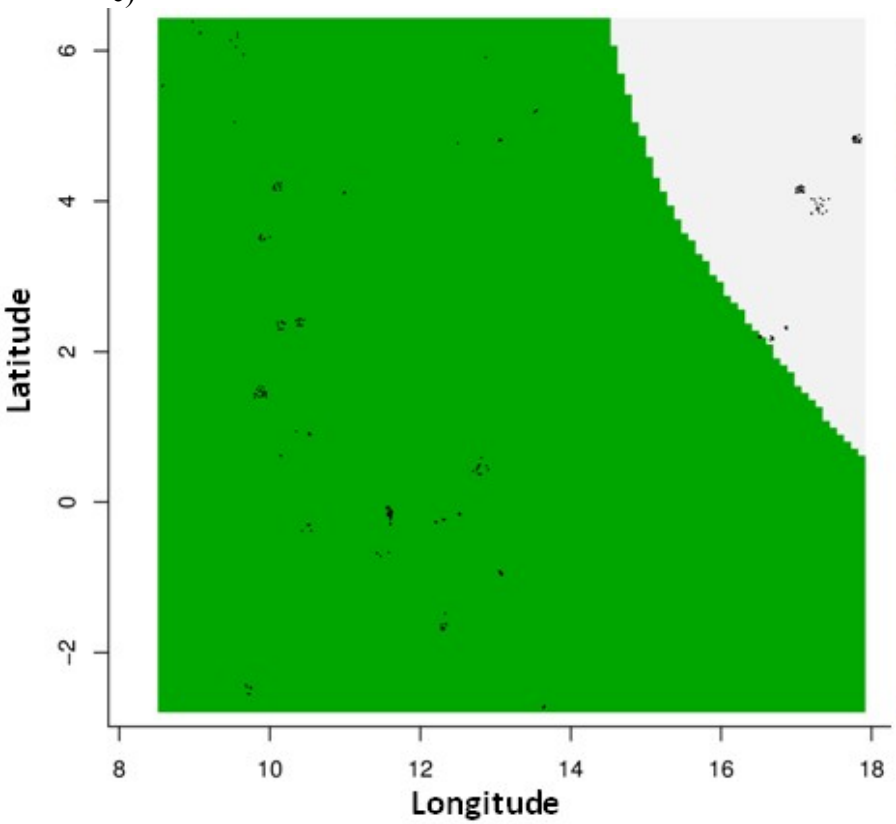

Nigeria
Central African Republic (CAR)
Oubangi
Sanaga
Cameroon
Congo
Sangha
Eq. Guinea
Ivindo
Republic of Congo
Gabon
Ogooué
Ngounié
Democratic Republic of Congo (DRC)
www.globalforestwatch.org

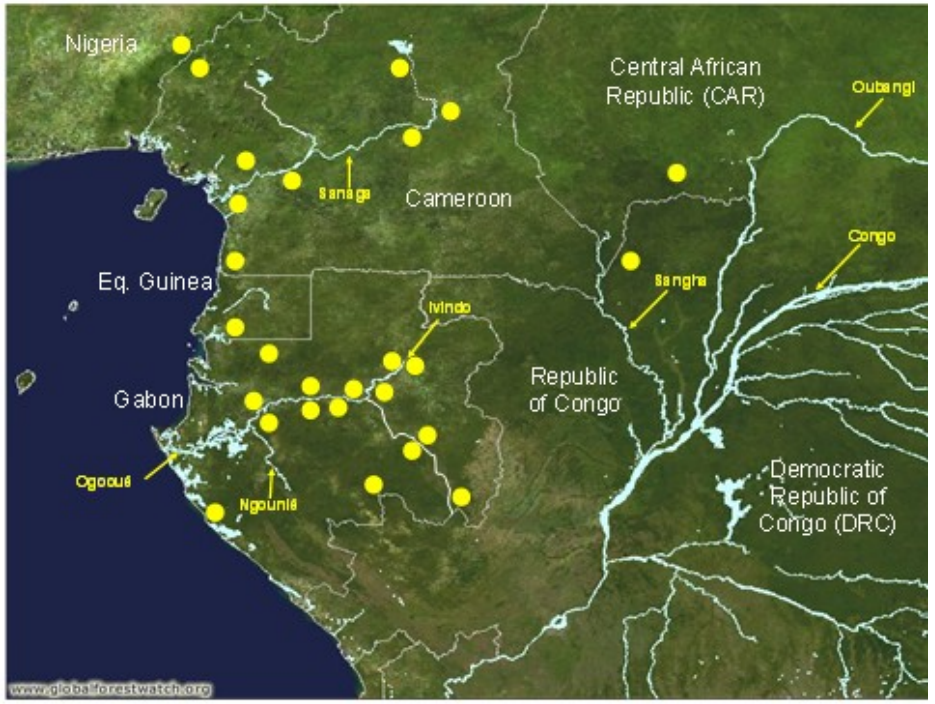
b)


Latitude
6
4
2
0
-2
8
10
12
14
16
Longitude


Nigeria
Central African Republic (CAR)
Oubangi
Sanaga
Cameroon
Congo
Sangha
Eq. Guinea
Mindo
Republic of Congo
Gabon
Ogooué
Ngounié
Democratic Republic of Congo (DRC)
www.globalforestwatch.org

c)


Latitude
6
4
2
0
−2
8
10
12
14
16
18
Longitude


Nigeria
Central African Republic (CAR)
Oubangi
Sanaga
Cameroon
Congo
Sangha
Eq. Guinea
Ivindo
Republic of Congo
Gabon
Ogooué
Ngounié
Democratic Republic of Congo (DRC)
www.globalforestwatch.org

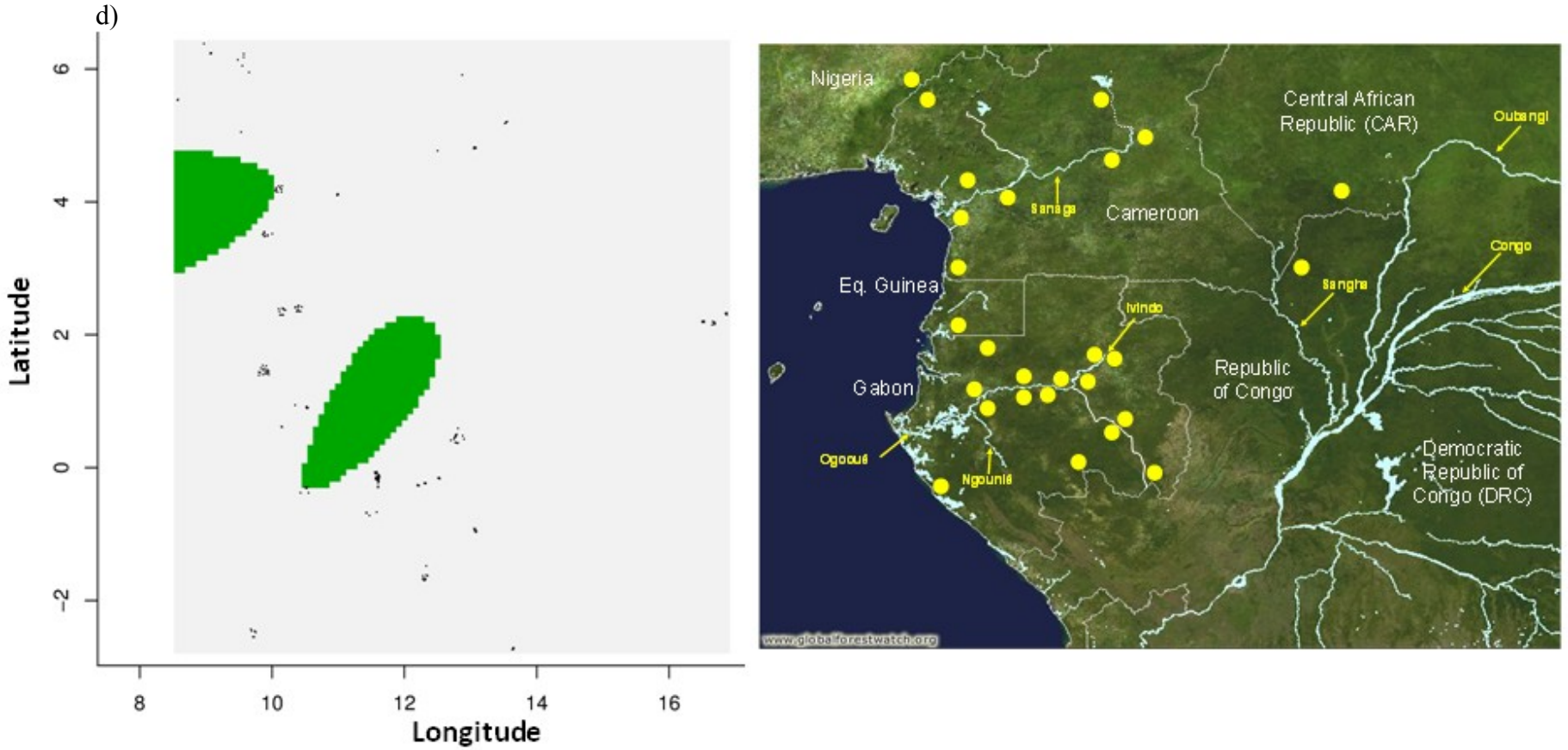


Figure 5. Maps showing sampling locations of (a) *C. dorsalis,* (b) *C. callipygus* and (c) *P. monticola* without first order relatives included. (d) *P. monticola* is also illustrated when NGO is excluded. For each species, the map on the left is a synthetic map of population membership of each K cluster using GENELAND and shows the ranges of the inferred genetic groups (green, white, and yellow). Black dots indicate sampling locations. For each species, the map on the right indicates the approximate locations (yellow dots) of sites represented on the left by black dots. Geographic distances between sites are not to scale in the map on the right.

Analyses of the entire *P. monticola* dataset with NGO (Site 39) samples gave two different estimates of K depending on whether first-order relatives were included or not. When first-order relatives are included, three clusters can be identified: cluster one comprises all NGO individuals (Site 39) in CAR and most of ND/CO (Site 38) samples (14 out of 20) from neighboring Republic of Congo. Cluster two comprises all GA (Site 35) samples in southwestern Gabon and cluster three comprises the rest (Supplementary Figures 6c). When first-order relatives are excluded, the number of clusters falls to two: one comprising all NGO (Site 39) and most of ND/CO (Site 38) samples (12 out of 17), while the other comprising the rest (Figure 5c).

Analyses of the same dataset without NGO (Site 39) samples and first-order relatives either included or excluded also gave two different numbers of clusters. When first-order relatives are included, the dataset fell into three clusters: cluster 1 comprised five out of 11 Ebo (Site 6) samples to the North of the Sanaga River and one MOO_South (Site 22) sample (MOO38) south of the Ogooué River. Cluster two comprised all GA (Site 35) samples and cluster 3 comprised the rest (Supplementary Figure 6c). When first-order relatives are removed, the number of clusters fell to two: cluster 1 contained two out of 10 Ebo (Site 6) samples (EBO20 and 35) and one MOO_South (Site 22) sample (MOO38) south of the Ogooué River, while cluster two comprises the rest (Figure 5d).

5.3.3.5. Test of Isolation by Distance using Isolation By Distance Web Service (IBDWS)

Mantel test indicated no significant correlation between genetic and geographic distances for C. dorsalis ($Z = 2583.2274$, $r = 0.0975$, $P= 0.2360$, $R^2 = 9.514e-03$) (Supplementary Figure 7a). The results remained similar after the pairwise geographic distances were log transformed ($Z = 14.9301$, $r = 0.1141$, $P= 0.2020$, $R^2 = 0.0130$) (Supplementary Figure 7a). In contrast, the *C. callipygus* Mantel test indicated a significant correlation between genetic and geographic distance ($Z = 4143.7771$, $r = 0.3477$, $R^2 = 0.121$, $P= 0.0020$) (Supplementary Figure 7b). A significant correlation for *C. callipygus* was also found after the geographic pairwise distances were log transformed ($Z = 21.6277$, $r = 0.2822$, $R^2 = 0.0796$, $P= 0.0040$) (Supplementary Figure 7b). *P. monticola* Mantel test indicates no significant correlation between genetic and geographic distance ($Z = 4610.4628$, $r = 0.1585$, $R^2 = 0.0251$, $P= 0.0570$) (Supplementary Figure 7c), but this association becomes significant after the geographic distance was log transformed ($Z = 24.0345$, $r = 0.1742$, $R^2 = 0.0303$, $P= 0.0260$) (Supplementary Figure 7c).

5.4. Discussion

The present study is the first to simultaneously assess patterns of mitochondrial and nuclear diversification among several species of duikers within the subfamily Cephalophinae in central Africa. Collectively results from the present study indicate that three areas within the study region appear to have played an important role in the diversification processes of forest duikers in central African rainforests: (1) the highlands of southwest Nigeria and adjacent southwest Cameroon, (2) the Sangha River basin, and (3) the Sanaga River and its associated basin.

Analyses of mitochondrial datasets from both *C. callipygus* and *P. monticola* identify the highlands of Nigeria and Cameroon as genetically distinct from the remaining sites southward of the study area (Figures 2b, 2c, 3c; Tables 3; Supplementary Table 3). Furthermore, most Nigerian samples that were initially identified as *P. monticola* with the

mitochondrial control region were later re-identified as belonging to the newly described species *P. walteri* using the mitochondrial cytochrome oxidase I (Colyn *et al.*, 2010). This new taxon was originally described from the Niger delta (Colyn *et al.*, 2010). So, findings of the present study extend its distribution as far as the Nigerian/Cameroonian border. The present study is not the first to highlight this region as a refugial center. Another study by Graham *et al.* (2005) assessed past and current ecological factors that may have affected species richness, endemism and turnover in the Gulf of Guinea highlands. Graham *et al.* (2005) found that areas with high and consistent annual rainfall showed the highest species and endemic richness, including three sites in southwest Cameroon: Mounts Cameroon, Manenguba, and Bakossi which have more than half of the endemic subspecies (Graham *et al.*, 2005). For instance, studies of the phylogeography of the murid rodent *Praomys misonnei* in the African tropics (Nicolas *et al.,* 2011) found that one sub-clade (Clade I2) was geographically restricted to the Korup National Park which is directly south of the Takamanda National park in Cameroon and East of the Cross River National Park in Nigeria. Similarly, Moodley and Bruford (2007) conducted a pan-African genetic analysis of the African bushbuck (*Tragelaphus scriptus*). Among the many region-specific haplogroups identified, they identified one ("Niger") whose geographic range spans east of Ghana to southwest Nigeria and adjacent southwest in Cameroon, supporting the genetic uniqueness of that same region. All these findings support the genetic uniqueness of biota of Southwest Nigeria and Cameroon (de Klerk *et al.*, 2002; Fishpool *et al.*, 2001) that could possibly be due to the combined effect of the Cross River as important biogeographical barrier (Nicolas *et al.*, 2011; Penner *et al.*, 2011) and the highlands of that area as putative forest refugia (Maley, 1996a). The present study only supports this area as an important center of biodiversity as evidenced in *P. monticola* and *C. callipygus*.

Second, the analysis of *C. callipygus* and *P. monticola* advocates a putative refugium in the Sangha basin as previously suggested (Anthony *et al.*, 2007). This finding was also supported by a significant signature of population expansion in this area for *C. callipygus* and *P. monticola* (Table 7). In addition, the analysis of the mitochondrial control region marker suggests that the Sangha basin might also constitutes an important center of *C. dorsalis* genetic diversity (Table 5). In fact, Anthony *et al.* (2007) reported a similar result for gorillas (*Gorilla gorilla*) in central Africa. They found that one haplotype (D2) was restricted to Dzanga-Sangha (within the same region). Anthony *et al.* (2007) explained their findings by the probable existence of a riparian refugium (Aide, Rivera, 1998; Colyn *et al.*, 1991; Leal, 2004; Meave *et al.*, 1991) at the confluence of several major central African rivers, including the Sangha and the Oubangi which eventually empty into the Congo River. The existence of riparian refugia has even led some authors to propose a distinct hypothesis which states that diversification processes in tropical lowland taxa may have occurred in major river watersheds during the arid phases of the Pleistocene (Aide, Rivera, 1998; Meave *et al.*, 1991). Other studies have already noted the high species diversity and endemicity of this area (Dowsett-Lemaire, Dowsett, 2000). Findings of the present study support the importance of the Sangha area as important center of diversity and putative refugia for central African taxa. Additionally, the present study has found a significant signature of population expansion at MM/MTA (Site 16) for *P. monticola* (Table 7) which may indicate a putative Pleistocene refugium as previously proposed (Anthony *et al.*, 2007; Born *et al.*, 2011; Muloko-Ntoutoume *et al.*, 2000; Sosef, 1994), but it could not be genetically differentiated from the remaining sites.

Third, the present study shows for the first time that the Sanaga river basin may have acted as both a fluvial refugium and/or a riverine barrier. Fluvial refugia have been previously proposed in the Sangha River basin for gorillas (*Gorilla gorilla*) (Anthony *et al.*, 2007) and in

the Zaire River basin (known as Congo River basin nowadays) for several species of simian primates (Colyn *et al.*, 1991). The latter study assessed primate radiations in the Zaire River basin using morphological characters (skin color and craniometrical measurements) and identified differentiated sets of taxa on opposite riverbanks (Colyn *et al.*, 1991). On the eastern bank of the Zaire-Lualaba River fourteen lowland taxa were identified: seven colobine, five guenons and two mangabeys. However, these lowland taxa were more closely related to taxa from the same riverbank than with counterparts that they have in the Central Rift montane forest in the east. When sister taxa from both lowland and highland areas come into contact, it was postulated that hybridization occurred, as is illustrated by *Colobus badius* which hybridizes extensively with all related subspecies in the region. A comparable picture of species richness of the same primates was also observed on the western bank of the Zaire River. However, several colobine subspecies appeared to have arisen in inter-fluvial blocks within riverbanks (Colyn *et al.*, 1991), probably due to the barrier effect of smaller tributary rivers of the Zaire. Nevertheless, Colyn *et al.* (1991) demonstrated that primate lineages within central Africa had two different center of origin: the central east Rift montane forest and the Zaire river Basin which may have been a fluvial refugium. It seems therefore that several central African river bodies may have acted as fluvial refugia, beside their traditional believed role of biogeographical barrier for some taxa. In the present study, the Sanaga River may have influenced the genetic structure of *P. monticola* and to a lesser extent for *C. dorsalis*. On one hand, mitochondrial analyses showed that *C. dorsalis* and *P. monticola* populations from the Sanaga basin and/or south Sanaga are differentiated from the remaining sites (Table 4, Supplementary Tables 2 and 4), while also showing significant signatures of population expansion (Tables 5 and 7). For instance, *P. monticola* AMOVA analysis supports a potential barrier effect for both Sanaga and Ogooué rivers since this group structure was also significant (F_{CT} = 12.64%, P = 0.019). However, the magnitude of this among group variance component was less than that observed for the Sanaga basin (Table 4). On the other hand, nuclear analyses of both *C. dorsalis* and *C. callipygus* suggest a clustering and/or barrier effect of the Sanaga basin and River, though with less support (Figures 4a, 5a, and 5b). For example, *C. callipygus* analysis shows that there is little support for a barrier effect of the Sanaga River owing to the fact that individuals from either bank fall into the same cluster (Figure 5b). However this result supports the genetic differentiation of sites from the Sanaga basin and northward.

In contrast to previous studies (Anthony *et al.*, 2007; Born *et al.*, 2011; Muloko-Ntoutoume *et al.*, 2000; Telfer *et al.*, 2003), our results show that the Ogooué River does not constitute on its own an effective barrier to the migration of forest duikers, despite an intensive sampling effort along its course from the source to the mouth. This lack of riverine barrier effect of the Ogooué may be attributable to its physical characteristics which could make it a less important as a barrier compared to the Sanaga River. The latter argument is not supported because the discharge, which is the average volume of water that flows per second in a given section of the water body, and the total length of the Ogooué (920 km and 4645 m^3/sec) are greater than those of the Sanaga (890 km, 3100 m^3/sec) (Vanden Bossche, Bernacsek, 1990). Other factors may then need to be considered such as the history, the seasonality, and possible narrower sections of the Ogooué River that may have permitted duikers to cross it more effectively. Nevertheless, *P. monticola* is genetically structured according to both Sanaga and Ogooué riverbanks. This result highlights a substantial regional substructuring among *P. monticola* populations. A recent study (Mboumba *et al.*, 2011) on the savannah specialist pygmy mice *Mus minutoides* in west-central Africa has also shown that the Ogooué is not an effective riverine barrier at least for some taxa. Mboumba *et al.* (2011) found that one mouse population (Mokekou) located north of the Ogooué River was more closely

related to distant peripheral populations (further south and east) than to three closely neighboring populations in Lopé National Park. In addition, the latter three closely neighboring mice populations were genetically more similar to one another than to Mokekou population, despite the fact that they were on opposite riverbanks. The present study also suggests that the Ogooué basin may have acted as fluvial refugium as illustrated in *C. callipygus* mitochondrial analysis (Table 6), which has never been proposed before.

This study illustrates differences in patterns of genetic structure between mitochondrial and nuclear microsatellite datasets. Mitochondrial control region is more suited for assessing historical patterns of genetic differentiation (Ntie *et al.*, 2010a) while nuclear microsatellites are more suited for the study of contemporary and fine scale evolutionary processes of central African duikers (Ntie *et al.*, 2010b). Mitochondrial markers have identified distinct genealogical lineages in Nigeria and Cameroon, the Sanaga River and/or its basin, and the Sangha basin, as well as signatures of demographic expansion at the regional level. In contrast, the nuclear data shows little structure aside from a potential division across the Sanaga River and a pronounced pattern of differentiation between the Sangha river basin and the rest of central Africa (Figure 5c). Additionally, including first order relatives does influence the outcome of the analysis as evidenced with GA (Site 35) samples which form their own cluster when included (Supplementary Table 6c), but do not when they are excluded (Figure 5c). Patterns of differentiation in the mitochondrial dataset may reflect the formation of refugial populations that were isolated during the drier and colder phases of the Pleistocene when forested habitats were restricted to highlands and lowland watersheds. When climate became warmer, duikers followed the expansion of the forest and expanded throughout central Africa, erasing any signatures of past population structure and leaving signatures of past population expansion. Alternatively, the greater mitochondrial structure evident in the present dataset might suggest that female duikers are more philopatric than males (Dubost, 1980). This observation combined with the fact that mitochondrial DNA is maternally inherited and is generally considered to have a lower genetic effective population size relative to nuclear DNA, might explain the greater genetic structure in mitochondrial as opposed to nuclear data.

Finally, there is substantial population genetic variation within each species, as evidenced in particular in *P. monticola* where several haplogroups are site or region specific. This suggests that a study carried out on a finer scale with a greater number of molecular markers could unravel more subtle patterns of population genetic structure. This can only be accomplished by sampling more intensively areas from which genetic material is still very limited, by using more powerful genetic marker such as Single Nucleotide Polymorphisms (SNPs), and high throughput molecular techniques such as Next Generation Sequencing (NGS). This will ultimately allow us to identify landscape features that may have been overlooked at the scope of this regional study.

Chapter 6: General Discussion

The rainforests of central Africa have attracted considerable attention from researchers, decision makers, and people inhabiting this region mainly because of their remarkable species richness and endemism (Brooks *et al.*, 2006; Linder, 2001; Mittermeier *et al.*, 1998; Reid, 1998), high economic value (e.g. logging, ecotourism) (Bouare, 2006; Gray, Ngolet, 1999; Hall *et al.*, 2003; Laurance *et al.*, 2006a; Laurance *et al.*, 2008; Ndoye, Tieguhong, 2004), and worldwide importance as a carbon sink in the regulation of greenhouse gases (Bombelli *et al.*, 2009; Ciais *et al.*). Therefore it appears urgent that more studies are carried out in this region to better understand how biological diversity came about and what strategies are needed to maintain it. This information is critical because it will increase our basic knowledge of factors driving evolutionary diversification in this region, help identify areas of high conservation value, and help predict the response of central African wildlife due to habitat degradation and climate change.

However, carrying out evolutionary research on cryptic mammalian taxa such as forest duikers requires geo-referenced non-invasive sampling strategies, such as it is the case in the present study. Indeed, adopting a non-invasive approach always leads to go through a few technical challenges prior to generating genomic data. It is important to run a pilot study to determine the best storage method and how long samples can be kept in storage media prior to DNA extraction. Previous studies have shown that mammalian fecal DNA degrades faster in the tropics due to high temperatures and humidity (Bayes *et al.*, 2000a; Wasser *et al.*, 1997), making work on tropical animals like duikers particularly challenging. Furthermore, there is little agreement in literature over the best storage method for artiodactyls or in fact any other animal feces (Frantzen *et al.*, 1998; Garnier *et al.*, 2001a; Johnson *et al.*, 2007; Muwanika *et al.*, 2007; Nsubuga *et al.*, 2004; Piggott, Taylor, 2003a; Vallet *et al.*, 2008; Van Hooft *et al.*, 2002). In order to address these challenges, a preliminary analysis was carried out on forest duiker feces stored under three standard storage media for 1-3 months duration. Findings from this study revealed that while silica beads are the best storage method for nuclear microsatellite analyses, the nucleic acid stabilizing medium RNA*later* preserves mitochondrial DNA for the longest period (Soto-Calderon *et al.*, 2009). In addition, extracting DNA as early as possible following collection always provides better DNA yields. Given that many of the sampling locations for phylogeographic studies of central African mammals are in extremely remote areas, storage in silica gel is recommended as it is cheap and easy to transport in the field. Moreover, as DNA degradation is likely to be most severe for single copy nuclear markers, it is also recommended that silica be the storage medium of choice. Further work should consider the possibility of developing on site DNA extraction protocols (Reed *et al.*, 1997; Tomlinson *et al.*, 2005) in order to minimize the degradation of DNA as much as possible prior to PCR amplification.

Another issue with studying central African forest mammals is that fecal samples for many rainforest artiodactyls cannot be readily distinguished (Bowkett *et al.*, 2009; van Vliet *et al.*, 2008) making species identification extremely difficult. To address this problem, the present study built a robust phylogenetic tree using reference samples from each duiker species (Ntie *et al.*, 2010a). Control region sequences from unknown duiker samples were then identified by looking at which species clade they fell (Ntie *et al.*, 2010a). This tree based identification method has numerous possible applications to the study of free-living central African artiodactyl species including fine-scale mapping of species distributions, identification of confiscated tissue in the regulation of bushmeat trade, and environmental impact assessments of human activities such as logging and hunting on wildlife populations (Wilkie *et al.*, 1992).

Notice that the utility of such a species identification methodology is only relevant when the whole geographic range of the studied taxa is covered, when the relationship between taxa within the group of interest is well established, and when the marker used has the appropriate genetic resolution and power to discriminate among selected species (Ntie *et al.*, 2010a).

Once the above technical difficulties for the use of duiker feces as source of DNA were resolved, a set of twelve nuclear microsatellites were put together into three multiplexes for fine scale analyses of contemporary population genetic structure. These microsatellite markers amplify across the focal taxa of this study, were highly polymorphic and powerful enough to distinguish individuals (Ntie *et al.*, 2010b). Such microsatellite multiplex assemblies should greatly facilitate studies of individual identification, parentage analysis, population size estimation and fine-scale analyses of population genetic structure in central African artiodactyls. However, with the advent of high throughput sequencing capability, future studies should make use of next generation sequencing to identify more nuclear microsatellite markers or single nucleotide polymorphisms (SNPs) in order to provide a much better level of resolution for detecting more subtle patterns of genetic differentiation among studied populations.

Lastly, findings from the present study were used to assess the impact of former Pleistocene refugia and riverine barriers on genetic structure of the three most abundant species in our sampling area; namely the bay duiker (*C. dorsalis*), the Peter's duiker (*C. callipygus*), and the blue duiker (*P. monticola*). We found that (1) southwestern Nigeria and southwestern Cameroon house genetically distinct populations of *C. callipygus* and *P. monticola*, (2) the Sanaga may constitute a partial riverine barrier to *C. callipygus* and *P. monticola* and (3) the Sangha River and potentially the Sanaga river basin may have acted as additional fluvial refugia for these species. Interestingly, mitochondrial haplotypes from Nigerian populations of *P. monticola* cluster with sequences from the newly described species *P. walteri*. This new taxon was originally described from the Niger delta (Colyn *et al.*, 2010). So, findings from this study extend its distribution as far as the Nigerian/Cameroonian border. Furthermore, strong patterns of demographic expansion in several highly differentiated haplogroups from all three taxa were observed at the regional level: Monte Alen and Monte Mitra, Sanaga basin, and Sangha basin. These patterns are consistent with a history of post-refugial expansion out of former Pleistocene forest refugia that may have existed during the dry phases of the Pleistocene and mirror a pattern of demographic expansion observed in one major haplogroup of western gorillas (Anthony *et al.*, 2007). However, there is little evidence in the nuclear dataset to support regionally differentiated haplogroups as has been observed in gorillas. Whereas post-refugial expansion in gorillas appears to have been limited, duiker population expansion may have been so extensive as to lead to a general pattern of population admixture [see (Johnson *et al.*, 2007)], dissimulating any past signatures of allopatric differentiation that may have occurred previously. Further comparative work is needed in order to verify these patterns across other multiple co-distributed species.

With regards to riverine effects, it is possible that central African rivers may have acted as riverine barriers, fluvial refugia, or both (Colyn *et al.*, 1991; Maley, 1989; Maley, 1996b; Maley, Brenac, 1998). One way to differentiate between these two hypotheses would be to see whether populations on opposite riverbanks are differentiated, and if not, whether genetic structure coincides with one or more major river basins. The present study shows that the Sanaga may have acted as both a partial riverine barrier (*C. callipygus* and *P. monticola*) and possibly as a fluvial refugium (*P. monticola*), and the Ogooué possibly also as a fluvial refugium (*C. callipygus*). In addition, the physical characteristics and history of the different

rivers in this study need to be considered in the light of such findings. For instance, a recent study (Sangen *et al.*, 2011) has found that between 13 - 22.4 thousand years BP the Sanaga river basin was characterized by recurrent channel migrations, floodplain reorganization and unstable vegetation (grasslands, woodlands, and gallery forests with bamboo thickets). These findings suggest that the Sanaga may have acted as a relatively weak riverine barrier. Future studies should then focus on filling in several gaps of our sampling area and extending the geographic scope of the present study to the river banks of the Congo River.

The present study has also opened up broad possibilities for research on many aspects of duiker ecology. For instance, one could develop a DNA barcoding approach for major plant groups in the region in order to study the diet of different duiker species. Another possible study could use multi-locus genotyping to estimate species abundance in a given area in response to different levels of hunting or habitat fragmentation (Mowry *et al.*, 2011; Poole *et al.*, 2011). Also, though feces have proven to be a reliable source of genetic material in the present study, they remain difficult to work with since about only half of our field collected samples were identified to species level and the time and cost to obtain reliable microsatellite data was high. Next Generation Sequencing (NGS) methods can now be used to obtain larger suites of genetic markers either through massive parallel sequencing of restriction sites to identify candidate SNPs (Davey, Blaxter, 2010) or targeted sequence capture of genetic markers from fecal samples (Perry *et al.*, 2010). The latter approach in particular could provide viable alternatives to non-invasive microsatellite genotyping and has already been demonstrated to work on primate fecal samples. These approaches could also provide the means of retrieving information on patterns of variation in genetic markers potentially under selection or associated with fitness or disease resistance, as has been recently developed and implemented in related domestic and wild bovid taxa (Cosart *et al.*, 2011). This kind of approach is especially attractive when multiple research groups collaborate and may foster international partnerships between African and US scientists, and their students interested in studying central African biodiversity.

Bibliography

Adams JM, Faure H (1997) Preliminary vegetation maps of the world since the last glacial maximum: An aid to archaeological understanding. *Journal of Archaeological Science* **24**, 623-647.

Adams JR, Waits LP (2007a) An efficient method for screening faecal DNA genotypes and detecting new individuals and hybrids in the red wolf (Canis rufus) experimental population area. *Conservation Genetics* **8**, 123-131.

Adams JR, Waits LP (2007b) An efficient method for screening faecal DNA genotypes and detecting new individuals and hybrids in the red wolf (Canis rufus) experimental population area. *Conservation Genetics* **8**, 123.

Aide TM, Rivera E (1998) Geographic patterns of genetic diversity in Poulsenia armata (Moraceae): implications for the theory of Pleistocene refugia and the importance of riparian forest. *Journal of Biogeography* **25**, 695-705.

Akaike H (1974) NEW LOOK AT STATISTICAL-MODEL IDENTIFICATION. *Ieee Transactions on Automatic Control* **AC19**, 716-723.

Albrechtsen L, Fa JE, Barry B, Macdonald DW (2005) Contrasts in availability and consumption of animal protein in Bioko Island, West Africa: the role of bushmeat. *Environmental Conservation* **32**, 340-348.

Amos W, Hoffman JI, Frodsham A*, et al.* (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes* **7**, 10-14.

Anhuf D, Ledru MP, Behling H*, et al.* (2006) Paleo-environmental change in Amazonian and African rainforest during the LGM. *Palaeogeography Palaeoclimatology Palaeoecology* **239**, 510-527.

Anthony NM, Johnson-Bawe M, Jeffery K*, et al.* (2007) The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20432-20436.

Ayres JM, Clutton-Brock TH (1992) River boundaries and species range size in Amazonian primates. *Am Nat* **140**, 531-537.

Baker CS, Lukoschek V, Lavery S*, et al.* (2006) Incomplete reporting of whale, dolphin and porpoise 'bycatch' revealed by molecular monitoring of Korean markets. *Animal Conservation* **9**, 474-482.

Baker CS, Palumbi SR (1994) WHICH WHALES ARE HUNTED - A MOLECULAR-GENETIC APPROACH TO MONITORING WHALING. *Science* **265**, 1538-1539.

Bakker EMV, Coetzee JA (1988) A Review of Late Quaternary Pollen Studies in East, Central and Southern-Africa. *Review of Palaeobotany and Palynology* **55**, 155-174.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37-48.

Banks SC, Skerratt LF, Taylor AC (2002) Female dispersal and relatedness structure in common wombats (Vombatus ursinus). *Journal of Zoology* **256**, 389-399.

Barendse W, Armitage SM, Kossarek LM*, et al.* (1994) A GENETIC-LINKAGE MAP OF THE BOVINE GENOME. *Nature Genetics* **6**, 227-235.

Barnes R, Greene K, Holland J, Lamm M (2002) Management and husbandry of duikers at the Los Angeles Zoo. *Zoo Biology* **21**, 107-121.

Barnes RFW (2002) The bushmeat boom and bust in West and Central Africa. *Oryx* **36**, 236-242.

Bayes MK, Smith KL, Alberts SC, Altmann J, Bruford MW (2000a) Testing the reliability of microsatellite typing from faecal DNA in the savannah baboon. *Conservation Genetics* **1**, 173-176.

Bayes MK, Smith KL, Alberts SC, Altmann J, Bruford MW (2000b) Testing the reliability of microsatellite typing from faecal DNA in the savannah baboon. *Conservation Genetics* **1**, 173.
Beh KJ, Riffkin CD, Davies KP, di Ienno K, Maddox JF (2000) Dinucleotide repeat polymorphism at the ovine McMA7, McMA10, McMA13, McMA16, McMA17, McMA27, McMA29, McMA42, McMA47 and McMA49 loci. *Animal Genetics* **31**, 228-229.
Beja-Pereira A, Zeyl E, Ouragh L*, et al.* (2004) Twenty polymorphic microsatellites in two of North Africa's most threatened ungulates: Gazella dorcas and Ammotragus lervia (Bovidae; Artiodactyla). *Molecular Ecology Notes* **4**, 452-455.
Bellis C, Ashton KJ, Freney L, Blair B, Griffiths LR (2003) A molecular genetic approach for forensic animal species identification. *Forensic Science International* **134**, 99-108.
Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution* **16**, 314-321.
Berger A, Bruschek M, Grethen C, Sperl W, Kofler B (2001) Poor storage and handling of tissue mimics mitochondrial DNA depletion. *Diagnostic Molecular Pathology* **10**, 55-59.
Bermingham E, Moritz C (1998) Comparative phylogeography: concepts and applications. *Molecular Ecology* **7**, 367-369.
Bidlack AL, Reed SE, Palsboll PJ, Getz WM (2007) Characterization of a western North American carnivore community using PCR-RFLP of cytochrome b obtained from fecal samples. *Conservation Genetics* **8**, 1511-1513.
Bishop MD, Kappes SM, Keele JW*, et al.* (1994) A Genetic-Linkage Map for Cattle. *Genetics* **136**, 619-639.
Bombelli A, Henry M, Castaldi S*, et al.* (2009) An outlook on the Sub-Saharan Africa carbon balance. *Biogeosciences* **6**, 2193-2205.
Bonnefille R, Hamilton AC, Linder HP, Riollet G (1990) 30,000-Year-Old Fossil Restionaceae Pollen from Central Equatorial Africa and Its Biogeographical Significance. *Journal of Biogeography* **17**, 307-314.
Bonnefille R, Riollet G, Buchet G*, et al.* (1995) Glacial/interglacial record from intertropical Africa, high resolution pollen and carbon data at Rusaka, Burundi. *Quaternary Science Reviews* **14**, 917-936.
Born C, Alvarez N, McKey D*, et al.* (2011) Insights into the biogeographical history of the Lower Guinea Forest Domain: evidence for the role of refugia in the intraspecific differentiation of Aucoumea klaineana. *Molecular Ecology* **20**, 131-142.
Bouare O (2006) A policy tool for establishing a balance between wildlife habitat preservation and the use of natural resources by rural people in South Africa. *African Journal of Ecology* **44**, 95-101.
Boucher J, Hanosset R, Augot D*, et al.* (2005a) Detection of Echinococcus multilocularis in wild boars in France using PCR techniques against larval form. **129**, 259-266.
Boucher JM, Hanosset R, Augot D*, et al.* (2005b) Detection of Echinococcus multilocularis in wild boars in France using PCR techniques against larval form. In: *Vet Parasitol*, pp. 259-266, Netherlands.
Bowkett AE, Plowman AB, Stevens JR, Davenport TRB, van Vuuren BJ (2009) Genetic testing of dung identification for antelope surveys in the Udzungwa Mountains, Tanzania. *Conservation Genetics* **10**, 251-255.
Bowkett AE, Rovero F, Marshall AR (2008) The use of camera-trap data to model habitat use by antelope species in the Udzungwa Mountain forests, Tanzania. *African Journal of Ecology* **46**, 479-487.

Bradley BJ, Chambers KE, Vigilant L (2001a) Accurate DNA-based sex identification of apes using non-invasive samples. *Conservation Genetics* **2**, 179-181.
Bradley BJ, Doran D, Robbins MM*, et al.* (2002) Comparative analyses of genetic social structure in wild gorillas (Gorilla gorilla) using DNA from feces and hair. *American Journal of Physical Anthropology*, 47-48.
Bradley BJ, Robbins MM, Williamson EA, Boesch C, Vigilant L (2001b) Assessment of reproductive strategies in wild mountain gorillas (Gorilla gorilla beringei) using DNA from feces. *American Journal of Physical Anthropology*, 43-43.
Brashares JS, Arcese P, Sam MK*, et al.* (2004) Bushmeat hunting, wildlife declines, and fish supply in West Africa. *Science* **306**, 1180-1183.
Bretagne S, Guillou J, Morand M, Houin R (1993a) Detection of Echinococcus multicolaris DNA in fox feces. *Parasitology* **106**, 193-199.
Bretagne S, Guillou JP, Morand M, Houin R (1993b) Detection of Echinococcus multilocularis DNA in fox faeces using DNA amplification. *Parasitology* **106 (Pt 2)**, 193-199.
Broderick D, Idaghdour Y, Korrida A, Hellmich J (2003) Gene flow in great bustard populations across the Strait of Gibraltar as elucidated from excremental PCR and mtDNA sequencing. *Conservation Genetics* **4**, 793-800.
Brooks TM, Mittermeier RA, da Fonseca GA*, et al.* (2006) Global biodiversity conservation priorities. *Science* **313**, 58-61.
Broquet T, Menard N, Petit E (2007a) Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics* **8**, 249.
Broquet T, Menard N, Petit E (2007b) Noninvasive population genetics: a review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. *Conservation Genetics* **8**, 249-260.
Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* **20**, 1004-1006, 1008-1010.
Brumfield RT, Capparella AP (1996) Historical diversification of birds in northwestern South America: A molecular perspective on the role of vicariant events. *Evolution* **50**, 1607-1624.
Buchanan FC, Crawford AM (1993) Ovine microsatellites at the OarFCB11, OarFCB128, OarFCB193, OarFCB266 and OarFCB304 loci. *Anim Genet* **24**, 145.
Burney CW, Brumfield RT (2009) Ecology Predicts Levels of Genetic Differentiation in Neotropical Birds. *American Naturalist* **174**, 358-368.
Bush GL (1975) MODES OF ANIMAL SPECIATION. *Annual Review of Ecology and Systematics* **6**, 339-364.
Bush MB (1994) AMAZONIAN SPECIATION - A NECESSARILY COMPLEX MODEL. *Journal of Biogeography* **21**, 5-17.
Bush MB, Piperno DR, Colinvaux PA*, et al.* (1992) A 14 300-Yr Paleoecological Profile of a Lowland Tropical Lake in Panama. *Ecological Monographs* **62**, 251-275.
Cardenas ML, Gosling WD, Sherlock SC*, et al.* (2011) The response of vegetation on the Andean flank in western Amazonia to Pleistocene climate change. *Science* **331**, 1055-1058.
Carnaval AC, Moritz C (2008) Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography* **35**, 1187-1201.
Carpaneto GM, Fusari A, Okongo H (2007) Subsistence hunting and exploitation of mammals in the Haut-Ogooud province, south-eastern Gabon. *Journal of Anthropological Sciences* **85**, 183-193.

Chandiwana S, Ornbjerg N (2003) Review of North-South and South-South cooperation and conditions necessary to sustain research capability in developing countries. *Journal of Health Population and Nutrition* **21**, 288-297.
Ciais P, Bombelli A, Williams M*, et al.* The carbon balance of Africa: synthesis of recent research studies. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences* **369**, 2038-2057.
Clifford SL, Anthony NM, Bawe-Johnson M*, et al.* (2004) Mitochondrial DNA phylogeography of western lowland gorillas (Gorilla gorilla gorilla). *Molecular Ecology* **13**, 1551-1565.
Colinvaux PA, De Oliveira PE, Bush MB (2000) Amazonian and neotropical plant communities on glacial time-scales: The failure of the aridity and refuge hypotheses. *Quaternary Science Reviews* **19**, 141-169.
Colyn M, Gautierhion A, Verheyen W (1991) A Reappraisal of Paleoenvironmental History in Central Africa - Evidence for a Major Fluvial Refuge in the Zaire Basin. *Journal of Biogeography* **18**, 403-407.
Colyn M, Hulselmans J, Sonet G*, et al.* (2010) Discovery of a new duiker species (Bovidae: Cephalophinae) from the Dahomey Gap, West Africa. *Zootaxa*, 1-30.
Constable JL, Ashley MV, Goodall J, Pusey AE (2001a) Noninvasive paternity assignment in Gombe chimpanzees. *Molecular Ecology* **10**, 1279-1300.
Constable JL, Ashley MV, Goodall J, Pusey AE (2001b) Noninvasive paternity assignment in Gombe chimpanzees. In: *Mol Ecol*, pp. 1279-1300, England.
Cosart T, Beja-Pereira A, Chen S*, et al.* (2011) Exome-wide DNA capture and next generation sequencing in domestic and wild species. In: *BMC Genomics*, p. 347, England.
Cosse M, Gonzalez S, Maldonado JE (2007) Cross-amplification tests of ungulate primers in the endangered Neotropical pampas deer (Ozotoceros bezoarticus). *Genetics and Molecular Research* **6**, 1118-1122.
Cracraft J, Prum RO (1988) PATTERNS AND PROCESSES OF DIVERSIFICATION - SPECIATION AND HISTORICAL CONGRUENCE IN SOME NEOTROPICAL BIRDS. *Evolution* **42**, 603-620.
Crawford AM, Dodds KG, Ede AJ*, et al.* (1995) An autosomal genetic linkage map of the sheep genome. *Genetics* **140**, 703-724.
Croes BM, Laurance WF, Lahm SA*, et al.* (2007) The influence of hunting on antipredator behavior in central African monkeys and duikers. *Biotropica* **39**, 257-263.
Crookes DJ, Ankudey N, Milner-Gulland EJ (2005) The value of a long-term bushmeat market dataset as an indicator of system dynamics. *Environmental Conservation* **32**, 333-339.
Da Silva JMC, Rylands AB, Da Fonseca GAB (2005) The fate of the Amazonian areas of endemism. *Conservation Biology* **19**, 689-694.
Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. In: *Brief Funct Genomics*, pp. 416-423, England.
de Klerk HM, Crowe TM, Fjeldsa J, Burgess ND (2002) Patterns of species richness and narrow endemism of terrestrial bird species in the Afrotropical region. *Journal of Zoology* **256**, 327-342.
de Merode E, Smith KH, Homewood K*, et al.* (2007) The impact of armed conflict on protected-area efficacy in Central Africa. *Biology Letters* **3**, 299-301.
Deagle BE, Tollit DJ (2007) Quantitative analysis of prey DNA in pinniped faeces: potential to estimate diet composition? *Conservation Genetics* **8**, 743.

Deagle BE, Tollit DJ, Jarman SN*, et al.* (2005) Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. In: *Mol Ecol*, pp. 1831-1842, England.
DeBusk GH (1998) A 37,500-year pollen record from Lake Malawi and implications for the biogeography of afromontane forests. *Journal of Biogeography* **25**, 479-500.
DeYoung RW, Honeycutt RL (2005) The molecular toolbox: Genetic techniques in wildlife ecology and management. *Journal of Wildlife Management* **69**, 1362-1384.
Dowsett-Lemaire F, Dowsett RJ (2000) Birds of the Lobeke Faunal Reserve, Cameroon, and its regional importance for conservation. *Bird Conservation International* **10**, 67-87.
Dreher BP, Winterstein SR, Scribner KT*, et al.* (2007) Noninvasive estimation of black bear abundance incorporating genotyping errors and harvested bear. *Journal of Wildlife Management* **71**, 2684-2693.
Dubost G (1979) Size of African Forest Artiodactyls as Determined by the Vegetation Structure. *African Journal of Ecology* **17**, 1-17.
Dubost G (1980) Ecology and Social-Behavior of the Blue Duiker - Small African Forest Ruminant. *Zeitschrift Fur Tierpsychologie-Journal of Comparative Ethology* **54**, 205-266.
Dubost G (1983) The Behavior of Cephalophus-Monticola Thunberg and Cephalophus-Dorsalis Gray, and Evolutive Situation of the Cephalophs among the Ruminants .2. *Mammalia* **47**, 281-310.
Dubost G (1984) Comparison of the Diets of Frugivorous Forest Ruminants of Gabon. *Journal of Mammalogy* **65**, 298-316.
Dubost G, Feer F (1992) Reproductive-Cycles of Small Ruminants in Northeastern Gabon as a Function of Food Resource Variability. *Mammalia* **56**, 25-43.
Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11**, 2571-2581.
Dupont LM, Jahns S, Marret F, Ning S (2000) Vegetation change in equatorial West Africa: time-slices for the last 150 ka. *Palaeogeography Palaeoclimatology Palaeoecology* **155**, 95-122.
Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.
Eaton MJ, Martin A, Thorbjarnarson J, Amato G (2009) Species-level diversification of African dwarf crocodiles (Genus Osteolaemus): A geographic and phylogenetic perspective. *Molecular Phylogenetics and Evolution* **50**, 496-506.
Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.
Ekrem T, Willassen E, Stur E (2007) A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution* **43**, 530-542.
Elenga H, Schwartz D, Vincens A (1994) Pollen Evidence of Late Quaternary Vegetation and Inferred Climate Changes in Congo. *Palaeogeography Palaeoclimatology Palaeoecology* **109**, 345-356.
Ellis AM, Bernard RTF (2005) Estimating the density of kudu (Tragelaphus strepsiceros) in subtropical thicket using line transect surveys of dung and DISTANCE software. *African Journal of Ecology* **43**, 362-368.
Emmons LH, Gautierhion A, Dubost G (1983) Community Structure of the Frugivorous Folivorous Forest Mammals of Gabon. *Journal of Zoology* **199**, 209-222.
Endler JA (1977) Geographic variation, speciation, and clines. *Monogr Popul Biol* **10**, 1-246.

Endler JA (1982) Pleistocene forest refuges: fact or fancy? In: *Biological diversification in the tropics* (ed. Prance GT), pp. 641-657. Columbia University Press, New York.
Eriksson J, Hohmann G, Boesch C, Vigilant L (2004) Rivers influence the population genetic structure of bonobos (Pan paniscus). *Mol Ecol* **13**, 3425-3435.
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.
Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1**, 47-50.
Fa JE, Albrechtsen L, Johnson PJ, Macdonald DW (2009) Linkages between household wealth, bushmeat and other animal protein consumption are not invariant: evidence from Rio Muni, Equatorial Guinea. *Animal Conservation* **12**, 599-610.
Fa JE, Juste J, Burn RW, Broad G (2002) Bushmeat consumption and preferences of two ethnic groups in Bioko Island, West Africa. *Human Ecology* **30**, 397-416.
Fa JE, Seymour S, Dupain J*, et al.* (2006) Getting to grips with the magnitude of exploitation: Bushmeat in the Cross-Sanaga Rivers region, Nigeria and Cameroon. *Biological Conservation* **129**, 497-510.
FAO (1999) Recognizing 'peste de petits ruminants'. A field manual. . Food and Agricultural Organization, United Nations, Rome, Italy.
Feer F (1989a) Comparaison des regimes alimentaires de Cephalophus callipygus et C. dorsalis, Bovides sympatriques de la foret sempervirente africaine. *Mammalia* **53**, 563-604.
Feer F (1989b) COMPARATIVE DIET OF CEPHALOPUS-CALLIPYGUS AND C-DORSALIS, SYMPATRIC BOVIDS OF THE AFRICAN SEMPERVIRENT FOREST. *Mammalia* **53**, 563-604.
Feer F (1989c) The Use of Space by 2 Sympatric Duikers (Cephalophus-Callipygus and Cephalophus-Dorsalis) in an African Rain-Forest - the Role of Activity Rhythms. *Revue D Ecologie-La Terre Et La Vie* **44**, 225-248.
Fernandez MH, Vrba ES (2005) A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biological Reviews* **80**, 269-302.
Fishpool LDC, Evans I, Royal Society for the Protection of B, Global Environment F, United Nations Development P (2001) *Important Bird Areas in Africa and Associated Islands: Priority Sites for Conservation* BirdLife International.
Fjeldsa J, Lovett JC (1997) Geographical patterns of old and young species in African forest biota: The significance of specific montane areas as evolutionary centres. *Biodiversity and Conservation* **6**, 325-346.
Focardi S, Isotti R, Tinelli A (2002) Line transect estimates of ungulate populations in a Mediterranean forest. *Journal of Wildlife Management* **66**, 48-58.
Fong JJ, Parham JF, Shi H, Stuart BL, Carter RL (2007) A genetic survey of heavily exploited, endangered turtles: caveats on the conservation value of trade animals. *Animal Conservation* **10**, 452-460.
Foran DR (2006) Relative degradation of nuclear and mitochondrial DNA: An experimental approach. *Journal of Forensic Sciences* **51**, 766-770.
Frantzen MAJ, Silk JB, Ferguson JWH, Wayne RK, Kohn MH (1998) Empirical evaluation of preservation methods for faecal DNA. *Molecular Ecology* **7**, 1423-1428.
Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925.
Gagneux P, Boesch C, Woodruff DS (1997) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* **6**, 861-868.

Garnier JN, Bruford MW, Goossens B (2001a) Mating system and reproductive skew in the black rhinoceros. *Molecular Ecology* **10**, 2031-2041.
Garnier JN, Bruford MW, Goossens B (2001b) Mating system and reproductive skew in the black rhinoceros. In: *Mol Ecol*, pp. 2031-2041, England.
Gautierhion A, Emmons LH, Dubost G (1980) A Comparison of the Diets of 3 Major Groups of Primary Consumers of Gabon (Primates, Squirrels and Ruminants). *Oecologia* **45**, 182-189.
Gerloff U, Schlotterer C, Rassmann K*, et al.* (1995) AMPLIFICATION OF HYPERVARIABLE SIMPLE SEQUENCE REPEATS (MICROSATELLITES) FROM EXCREMENTAL DNA OF WILD LIVING BONOBOS (PAN-PANISCUS). *Molecular Ecology* **4**, 515-518.
Gibson UE, Heid CA, Williams PM (1996) A novel method for real time quantitative RT-PCR. *Genome Res.* **6**, 995-1001.
Graham CH, Smith TB, Languy M (2005) Current and historical factors influencing patterns of species richness and turnover of birds in the Gulf of Guinea highlands. *Journal of Biogeography* **32**, 1371-1384.
Gray C, Ngolet F (1999) Lambarene, okoume and the transformation of labor along the Middle Ogooue (Gabon), 1870-1945. *Journal of African History* **40**, 87-107.
Greenwood AD, Paabo S (1999) Nuclear insertion sequences of mitochondrial DNA predominate in hair but not in blood of elephants. *Molecular Ecology* **8**, 133-137.
Guillot G, Mortier F, Estoup A (2005) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes* **5**, 712-715.
Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**, 361-372.
Haberle SG (2005) A 23,000-yr pollen record from Lake Euramoo, Wet Tropics of NE Queensland, Australia. *Quaternary Research* **64**, 343-356.
Haffer J (1969) Speciation in amazonian forest birds. *Science* **165**, 131-137.
Haffer J (1997) Alternative models of vertebrate speciation in Amazonia: An overview. *Biodiversity and Conservation* **6**, 451-476.
Haffer J, Prance GT (2001) Climatic forcing of evolution in Amazonia during the Cenozoic: On the refuge theory of biotic differentiation. *Amazoniana-Limnologia Et Oecologia Regionalis Systemae Fluminis Amazonas* **16**, 579-605.
Hall JS, Harris DJ, Medjibe V, Ashton PMS (2003) The effects of selective logging on forest structure and tree species composition in a Central African forest: implications for management of conservation areas. *Forest Ecology and Management* **183**, 249-264.
Hammond RL, Macasero W, Flores B*, et al.* (2001) Phylogenetic reanalysis of the Saudi gazelle and its implications for conservation. *Conservation Biology* **15**, 1123-1133.
Hasegawa M, Kishino H, Yano TA (1985) DATING OF THE HUMAN APE SPLITTING BY A MOLECULAR CLOCK OF MITOCHONDRIAL-DNA. *Journal of Molecular Evolution* **22**, 160-174.
Hayes FE, Sewlal JAN (2004) The Amazon River as a dispersal barrier to passerine birds: effects of river width, habitat and taxonomy. *Journal of Biogeography* **31**, 1809-1818.
Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences* **270**, 313-321.
Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. *Plos Biology* **2**, 1657-1663.
Hedmark E, Ellegren H (2007) DNA-based monitoring of two newly founded Scandinavian wolverine populations. *Conservation Genetics* **8**, 843-852.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.
Heydon MJ, Bulloh P (1997) Mousedeer densities in a tropical rainforest: The impact of selective logging. *Journal of Applied Ecology* **34**, 484-496.
Heymans J, Lejoly J (1981) Sur la répartition des Cephalophinae en forêt ombrophile (Région du Haut-Zaïre, Rép. du Zaïre) (Bovidae). *Revue de Zoologie Africaine* **95**, 1-10.
Hickerson MJ, Meyer CP, Moritz C (2006) DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology* **55**, 729-739.
Hoelzel AR, Hancock JM, Dover GA (1991) Evolution of the Cetacean Mitochondrial D-Loop Region. *Molecular Biology and Evolution* **8**, 475-493.
Hofmann T, Roth H (2003) Feeding preferences of duiker (Cephalophus maxwelli, C-rufilatus, and C-niger) in Ivory Coast and Ghana. *Mammalian Biology* **68**, 65-77.
Hoss M, Kohn M, Paabo S, Knauer F, Schroder W (1992) Excrement analysis by PCR. *Nature* **359**, 199.
Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.
Hung C, Li S, Lee L (2004a) Faecal DNA typing to determine the abundance and spatial organisation of otters (Lutra lutra) along two stream systems in Kinmen. *Animal Conservation* **7**, 301–311.
Hung CM, Li SH, Lee LL (2004b) Faecal DNA typing to determine the abundance and spatial organisation of otters (Lutra lutra) along two stream systems in Kinmen. *Animal Conservation* **7**, 301-311.
Idaghdour Y, Broderick D, Korrida A (2003) Faeces as a source of DNA for molecular studies in a threatened population of great bustards. *Conservation Genetics* **4**, 789-792.
Jahns S, Huls M, Sarnthein M (1998) Vegetation and climate history of west equatorial Africa based on a marine pollen record off Liberia (site GIK 16776) covering the last 400,000 years. *Review of Palaeobotany and Palynology* **102**, 277-288.
Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**, 1801-1806.
Jalil MF, Cable J, Inyor JS*, et al.* (2008) Riverine effects on mitochondrial structure of Bornean orang-utans (Pongo pygmaeus) at two spatial scales. *Molecular Ecology* **17**, 2898-2909.
Jansen JHF, Alderliesten C, Houston CM*, et al.* (1989) Aridity in Equatorial Africa during the Last 225,000 Years - a Record of Opal Phytoliths Fresh-Water Diatoms from the Zaire (Congo) Deep-Sea Fan (Northeast Angola Basin). *Radiocarbon* **31**, 557-569.
Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *Bmc Genetics* **6**.
Johnson MB, Clifford SL, Goossens B*, et al.* (2007) Complex phylogeographic history of central African forest elephants and its implications for taxonomy. *Bmc Evolutionary Biology* **7**.
Johnston AR, Morikawa MK, Ntie S, Anthony NM (2011) Evaluating DNA barcoding criteria using African duiker antelope (Cephalophinae) as a test case. *Conservation Genetics* **12**, 1173-1182.
Jolly D, Harrison SP, Damnati B, Bonnefille R (1998) Simulated climate and biomes of Africa during the late quaternary: Comparison with pollen and lake status data. *Quaternary Science Reviews* **17**, 629-657.

Jolly D, Taylor D, Marchant R*, et al.* (1997) Vegetation dynamics in central Africa since 18,000 yr BP: pollen records from the interlacustrine highlands of Burundi, Rwanda and western Uganda. *Journal of Biogeography* **24**, 495-512.
Joseph L, Moritz C, Hugall A (1995) Molecular Support for Vicariance as a Source of Diversity in Rain-Forest. *Proceedings of the Royal Society of London Series B-Biological Sciences* **260**, 177-182.
Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed. Munro HN), pp. 21-132. Academic Press, New York.
Kanyamibwa S (1998) Impact of war on conservation: Rwandan environment and wildlife in agony. *Biodiversity and Conservation* **7**, 1399-1406.
Karesh WB, Hart JA, Hart TB*, et al.* (1995) Health evaluation of five sympatric duiker species (Cephalophus spp). *Journal of Zoo and Wildlife Medicine* **26**, 485-502.
Kaukinen J, Varvio SL (1993) Eight polymorphic bovine microsatellites. *Animal Genetics* **24**, 148-148.
Kershaw AP, Bretherton SC, van der Kaars S (2007) A complete pollen record of the last 230 ka from Lynch's Crater, north-eastem Australia. *Palaeogeography Palaeoclimatology Palaeoecology* **251**, 23-45.
Kim KS, Min MS, An JH, Lee H (2004) Cross-species amplification of Bovidae microsatellites and low diversity of the endangered Korean goral. *Journal of Heredity* **95**, 521-525.
Kim SY, Scourse J, Marret F, Lim DI (2010) A 26,000-year integrated record of marine and terrestrial environmental change off Gabon, west equatorial Africa. *Palaeogeography Palaeoclimatology Palaeoecology* **297**, 428-438.
Kingdon J (1997) *The Kingdon field guide to African mammals* Academic Press, London and New York.
Kirkpatrick JB, Fowler M (1998) Locating likely glacial forest refugia in Tasmania using palynological and ecological information to test alternative climatic models. *Biological Conservation* **85**, 171-182.
Kogi J, Yeh CC, Bhebhe E*, et al.* (1995) Caprine microsatellite dinucleotide repeat polymorphisms at the SR-CRSP-11, SR-CRSP-12, SR-CRSP-13, SR-CRSP-14 and SR-CRSP-15 loci. *Anim Genet* **26**, 449.
Kohn M, Knauer F, Stoffella A, Schröder W, Pääbo S (1995) Conservation genetics of the European Brown bear - A study using excremental PCR of nuclear and mitochondrial sequences. *Molecular Ecology* **4**, 95-103.
Kohn MH, Wayne RK (1997a) Facts from feces revisited. *Trends in Ecology & Evolution* **12**, 223-227.
Kohn MH, Wayne RK (1997b) Facts from feces revisited. *Trends in Ecology and Evolution* **12**, 223.
Kohn MH, York EC, Kamradt DA*, et al.* (1999a) Estimating population size by genotyping faeces. *Proceedings of the Royal Society Biological Sciences Series B* **266**, 657.
Kohn MH, York EC, Kamradt DA*, et al.* (1999b) Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London Series B-Biological Sciences* **266**, 657-663.
Konovalov DA, Manning C, Henshaw MT (2004) KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. *Molecular Ecology Notes* **4**, 779-782.
Kranz KR, Lumpkin S (1982) Notes on the yellow-backed duiker Cephalophus sylvicultor in captivity with comments on its natural history. *International Zoo Yearbook* **22**, 232-240.

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5**, 150-163.
Lahm SA (1993a) *Ecology and Economics of Human/Wildlife Interaction in Northeastern Gabon* Doctorate, New York University.
Lahm SA (1993b) Utilization of forest resources and local variation of wildlife populations in northeastern Gabon. *Man and the Biosphere Series* **13**, 213-226.
Lahm SA (1993c) 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* (ed. Hladik CM, Hladik, A., Linares, O.F., Pagezy, H., Semple A., Hadley, M.), pp. 213-227. UNESCO, Paris.
Lahm SA, Kombila M, Swanepoel R, Barnes RFW (2007) Morbidity and mortality of wild animals in relation to outbreaks of Ebola haemorrhagic fever in Gabon, 1994-2003. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **101**, 64-78.
Lahm SA, 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* (ed. History NMoN), pp. 383-417, Washington D.C.
Lantuéjoul C (2002) *Geostatistical Simulation: Models and Algorithms* Springer.
Laurance WF, Alonso A, Lee M, Campbell P (2006a) Challenges for forest conservation in Gabon, central Africa. *Futures* **38**, 454-470.
Laurance WF, Croes BM, Tchignoumba L*, et al.* (2006b) Impacts of roads and hunting on central African rainforest mammals. *Conservation Biology* **20**, 1251-1261.
Laurance WF, Fay JM, Parnell RJ*, et al.* (2008) Does rainforest logging threaten marine turtles? *Oryx* **42**, 246-251.
Leache AD, Reeder TW (2002) Molecular systematics of the Eastern Fence Lizard (Sceloporus undulatus): A comparison of parsimony, likelihood, and Bayesian approaches. *Systematic Biology* **51**, 44-68.
Leal ME (2004) *The African rain forest during the Last Glacial Maximum, an archipelago of forests in a sea of grass*, Wageningen University.
Lehman SM (2004) Distribution and diversity of primates in Guyana: Species-area relationships and riverine barriers. *International Journal of Primatology* **25**, 73-95.
Leroy SAG, Dupont LM (1997) Marine palynology of the ODP site 658 (N-W Africa) and its contribution to the stratigraphy of Late Pliocene. *Geobios* **30**, 351-359.
Lessa EP, Cook JA, Patton JL (2003) Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10331-10334.
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.
Linder HP (2001) Plant diversity and endemism in sub-Saharan tropical Africa. *Journal of Biogeography* **28**, 169-182.
Lorenz JG, Jackson WE, Beck JC, Hanner R (2005) The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philosophical Transactions of the Royal Society B-Biological Sciences* **360**, 1869-1877.
Lorenzini R, Posillico M, Lovari S, Petrella A (2004) Non-invasive genotyping of the endangered Apennine brown bear: a case study not to let one's hair down. *Animal Conservation* **7**, 199-209.
Lumpkin S, Kranz KR (1984) Cephalophus sylvicultor. *Mammalian Species*, 1-7.
Maley J (1989) Late quaternary climatic changes in the African rain forest: forest refugia and the major role of sea surface temperature variations. *Paleoclimatology and*

paleometeorology: modern and past patterns of global atmospheric transport. Proceedings of a NATO Advanced Research Workshop held at Oracle, AZ, USA, 17-19 November 1987., 585-616.
Maley J (1996a) The African rainforest - main characteristics of changes in vegetation and climate change from the Upper Cretaceous to the Quaternary. *Proceedings of the Royal Society of Edinburgh* **104B**, 31-73.
Maley J (1996b) Le cadre paléoenvironnementa1 des refuges forestiers africains: quelques données et hypothèses *The Biodiversity of African Piants*, 519-535.
Maley J, Brenac P (1998) Vegetation dynamics, palaeoenvironments and climatic changes in the forests of western Cameroon during the last 28,000 years BP. *Review of Palaeobotany and Palynology* **99**, 157-187.
Marchant R, Taylor D, Hamilton A (1997) Late Pleistocene and Holocene history at Mubwindi Swamp, southwest Uganda. *Quaternary Research* **47**, 316-328.
Marret F, Scourse J, Jansen JHF, Schneider R (1999) Climate and palaeoceanographic changes in west Central Africa during the last deglaciation: palynological investigation. *Comptes Rendus De L Academie Des Sciences Serie Ii Fascicule a-Sciences De La Terre Et Des Planetes* **329**, 721-726.
Maudet C, Luikart G, Taberlet P (2001) Development of microsatellite multiplexes for wild goats using primers designed from domestic Bovidae. *Genetics Selection Evolution* **33**, S193-S203.
Mboumba JF, Deleporte P, Colyn M, Nicolas V (2011) Phylogeography of Mus (Nannomys) minutoides (Rodentia, Muridae) in West Central African savannahs: singular vicariance in neighbouring populations. *Journal of Zoological Systematics and Evolutionary Research* **49**, 77-85.
Meave J, Kellman M, Macdougall A, Rosales J (1991) Riparian Habitats as Tropical Forest Refugia. *Global Ecology and Biogeography Letters* **1**, 69-76.
Mercader J, Runge F, Vrydaghs L*, et al.* (2000) Phytoliths from archaeological sites in the tropical forest of Ituri, Democratic Republic of Congo. *Quaternary Research* **54**, 102-112.
Minnemeyer S, Walker T, Collomb J-G, Cotton L, Bryant D (2002) An Analysis of Access into Central Africa's Rainforests (ed. Holmes K). World Resources Institute, Washington, DC.
Mittermeier RA, Myers N, Thomsen JB, da Fonseca GAB, Olivieri S (1998) Biodiversity hotspots and major tropical wilderness areas: Approaches to setting conservation priorities. *Conservation Biology* **12**, 516-520.
Mockrin MH (2010) Duiker demography and dispersal under hunting in Northern Congo. *African Journal of Ecology* **48**, 239-247.
Mommens G, Coppieters W, Vandeweghe A, Vanzeveren A, Bouquet Y (1994) DINUCLEOTIDE REPEAT POLYMORPHISM AT THE BOVINE MM12E6 AND MM8D3 LOCI. *Animal Genetics* **25**, 368-368.
Moore MK, Bemiss JA, Rice SM, Quattro JM, Woodley CM (2003) Use of restriction fragment length polymorphisms to identify sea turtle eggs and cooked meats to species. *Conservation Genetics* **4**, 95-103.
Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (Pan troglodytes verus). *Molecular Ecology* **10**, 1835-1844.
Moritz C, Cicero C (2004) DNA barcoding: Promise and pitfalls. *Plos Biology* **2**, 1529-1531.

Moritz C, Patton JL, Schneider CJ, Smith TB (2000) Diversification of rainforest faunas: An integrated molecular approach. *Annual Review of Ecology and Systematics* **31**, 533-563.
Mowry RA, Gompper ME, Beringer J, Eggert LS (2011) River Otter Population Size Estimation Using Noninvasive Latrine Surveys. *Journal of Wildlife Management* **75**, 1625-1636.
Muchaal PK, Ngandjui G (1999) Impact of village hunting on wildlife populations in the western Dia Reserve, Cameroon. *Conservation Biology* **13**, 385-396.
Muloko-Ntoutoume N, Petit RJ, White L, Abernethy K (2000) Chloroplast DNA variation in a rainforest tree (Aucoumea klaineana, Burseraceae) in Gabon. *Molecular Ecology* **9**, 359-363.
Murphy MA, Waits LP, Kendall C (2000) Quantitative evaluation of fecal drying methods for brown bear DNA analysis. *Wildlife Society Bulletin* **28**, 951-957.
Muwanika VB, Nyakaana S, Siegismund HR, Arctander P (2007) Population genetic structure of the common warthog (Phacochoerus africanus) in Uganda: evidence for a strong philopatry among warthogs and social structure breakdown in a disturbed population. *African Journal of Ecology* **45**, 22-30.
Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.
Ndoye O, Tieguhong JC (2004) Forest resources and rural livelihoods: The conflict between timber and non-timber forest products in the Congo Basin. *Scandinavian Journal of Forest Research* **19**, 36-44.
Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. *Genetics* **105**, 207-217.
Newing H (2001) Bushmeat hunting and management: implications of duiker ecology and interspecific competition. *Biodiversity and Conservation* **10**, 99-118.
Nichol JE (1999) Geomorphological evidence and Pleistocene refugia in Africa. *Geographical Journal* **165**, 79-89.
Nicolas V, Mboumba JF, Verheyen E*, et al.* (2008) Phylogeographic structure and regional history of Lemniscomys striatus (Rodentia: Muridae) in tropical Africa. *Journal of Biogeography* **35**, 2074-2089.
Nicolas V, Missoup AD, Denys C*, et al.* (2011) The roles of rivers and Pleistocene refugia in shaping genetic diversity in Praomys misonnei in tropical Africa. *Journal of Biogeography* **38**, 191-207.
Nicolas V, Querouil S, Verheyen E*, et al.* (2006) Mitochondrial phylogeny of African wood mice, genus Hylomyscus (Rodentia, Muridae): Implications for their taxonomy and biogeography. *Molecular Phylogenetics and Evolution* **38**, 779-793.
Nielsen MR (2006) Importance, cause and effect of bushmeat hunting in the Udzungwa Mountains, Tanzania: Implications for community based wildlife management. *Biological Conservation* **128**, 509-516.
Noss AJ (1999) Censusing rainforest game species with communal net hunts. *African Journal of Ecology* **37**, 1-11.
Nsubuga AM, Robbins MM, Roeder AD*, et al.* (2004) Factors affecting the amount of genomic DNA extracted from ape faeces and the identification of an improved sample storage method. *Molecular Ecology* **13**, 2089-2094.
Ntie S, Johnston AR, Mickala P*, et al.* (2010a) A molecular diagnostic for identifying central African forest artiodactyls from faecal pellets. *Animal Conservation* **13**, 80-93.
Ntie S, Soto-Calderón ID, Eaton MJ, Anthony NM (2010b) Cross-species amplification of bovid microsatellites in central African duikers (genus Cephalophus) and other sympatric artiodactyls. *Molecular Ecology Resources* **10**, 1059 - 1065.

Okouyi JO, Posso P, Lepretre A, Scaps P (2002) Estimation of large mammal densities in the Ipassa-Makokou reserve (Gabon). *Bulletin De La Societe Zoologique De France* **127**, 121-135.
Oliveira R, Castro D, Godinho R, Luikart G, Alves PC (2010) Species identification using a small nuclear gene fragment: application to sympatric wild carnivores from South-western Europe. *Conservation Genetics* **11**, 1023-1032.
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.
Paabo S, Wilson AC (1988) Polymerase Chain-Reaction Reveals Cloning Artifacts. *Nature* **334**, 387-388.
Pages M, Maudet C, Bellemain E*, et al.* (2009) A system for sex determination from degraded DNA: a useful tool for palaeogenetics and conservation genetics of ursids. *Conservation Genetics* **10**, 897-907.
Patton JL, da Silva MNF, J.R. M (1998) Rivers, refuges and ridges. In: *Endless forms: species and speciation* (eds. Howard DJ, Berlocher SH), pp. 202-216. Oxford University Press, New York.
Patton JL, Dasilva MNF, Malcolm JR (1994) Gene Genealogy and Differentiation among Arboreal Spiny Rats (Rodentia, Echimyidae) of the Amazon Basin - a Test of the Riverine Barrier Hypothesis. *Evolution* **48**, 1314-1323.
Penner J, Wegmann M, Hillers A, Schmidt M, Rodel MO (2011) A hotspot revisited - a biogeographical analysis of West African amphibians. *Diversity and Distributions* **17**, 1077-1088.
Penty JM, Henry HM, Ede AJ, Crawford AM (1993) Ovine microsatellites at the OarAE16, OarAE54, OarAE57, OarAE119 and OarAE129 loci. *Anim Genet* **24**, 219.
Perry GH, Marioni JC, Melsted P, Gilad Y (2010) Genomic-scale capture and sequencing of endogenous DNA from feces. *Mol Ecol* **19**, 5332-5344.
Piggott MP, Taylor AC (2003a) Extensive evaluation of faecal preservation and DNA extraction methods in Australian native and introduced species. *Australian Journal of Zoology* **51**, 341-355.
Piggott MP, Taylor AC (2003b) Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. *Wildlife Research* **30**, 1-13.
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. *Journal of Zoology* **271**, 140-147.
Pintaud JC, Jaffre T, Puig H (2001) Chorology of New Caledonian palms and possible evidence of Pleistocene rain forest refugia. *Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences* **324**, 453-463.
Plumptre AJ, Bizumuremyi JB, Uwimana F, Ndaruhebeye JD (1997) The effects of the Rwandan civil war on poaching of ungulates in the Parc National des Volcans. *Oryx* **31**, 265-273.
Poole KG, Reynolds DM, Mowat G, Paetkau D (2011) Estimating Mountain Goat Abundance Using DNA From Fecal Pellets. *Journal of Wildlife Management* **75**, 1527-1534.
Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.
Poulsen JR, Clark CJ, Mavah G, Elkan PW (2009) Bushmeat Supply and Consumption in a Tropical Logging Concession in Northern Congo. *Conservation Biology* **23**, 1597-1608.

Prentice IC, Jolly D, participants B (2000) Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. *Journal of Biogeography* **27**, 507-519.
Prigioni C, Remonti L, Balestrieri A*, et al.* (2006) Estimation of european otter (Lutra lutra) population size by fecal DNA typing in southern Italy. *Journal of Mammalogy* **87**, 855-858.
Prins HHT, de Boer WF, van Oeveren H*, et al.* (2006) Co-existence and niche segregation of three small bovid species in southern Mozambique. *African Journal of Ecology* **44**, 186-198.
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
Rambaut A, Drummond AJ (2007) TRACER.
Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* **19**, 2092-2100.
Reed JZ, Tollit DJ, Thompson PM, Amos W (1997) Molecular scatology: the use of molecular genetic analysis to assign species, sex and individual identity to seal faeces. *Mol Ecol* **6**, 225-234.
Reid WV (1998) Biodiversity hotspots. *Trends in Ecology & Evolution* **13**, 275-280.
Rizkalla C, Blanco-Silva F, Gruver S (2007) Modeling the impact of Ebola and bushmeat hunting on western lowland gorillas. *Ecohealth* **4**, 151-155.
Robinson TJ, Wilson V, Gallagher DS, Jr.*, et al.* (1996) Chromosomal evolution in duiker antelope (Cephalophinae: Bovidae): karyotype comparisons, fluorescence in situ hybridization, and rampant X chromosome variation. *Cytogenet Cell Genet* **73**, 116-122.
Rosen DE (1978) VICARIANT PATTERNS AND HISTORICAL EXPLANATION IN BIOGEOGRAPHY. *Systematic Zoology* **27**, 159-188.
Ross HA, Lento GM, Dalebout ML*, et al.* (2003) DNA Surveillance: Web-based molecular identification of whales, dolphins, and porpoises. *Journal of Heredity* **94**, 111-114.
Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.
Saiki RK, Gelfand DH, Stoffel S*, et al.* (1988) Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA-Polymerase. *Science* **239**, 487-491.
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425.
Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual* Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.
Sangen M (2011) New results on palaeoenvironmental conditions in equatorial Africa derived from alluvial sediments of Cameroonian rivers. *Proceedings of the Geologists Association* **122**, 212-223.
Sangen M, Neumann K, Eisenberg J (2011) Climate-induced fluvial dynamics in tropical Africa around the last glacial maximum? *Quaternary Research* **76**, 417-429.
Schneider C, Moritz C (1999) Rainforest refugia and evolution in Australia's Wet Tropics. *Proceedings of the Royal Society of London Series B-Biological Sciences* **266**, 191-196.
Schneider C, Roessli D, Excoffier L (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva, Geneva.
Schneider CJ, Cunningham M, Moritz C (1998) Comparative phylogeography and the history of endemic vertebrates in the Wet Tropics rainforests of Australia. *Molecular Ecology* **7**, 487-498.

Servant M, Maley J, Turcq B*, et al.* (1993) TROPICAL FOREST CHANGES DURING THE LATE QUATERNARY IN AFRICAN AND SOUTH-AMERICAN LOWLANDS. *Global and Planetary Change* **7**, 25-40.
Shields GF, Kocher TD (1991) Phylogenetic-Relationships of North-American Ursids Based on Analysis of Mitochondrial-DNA. *Evolution* **45**, 218-221.
Slatkin M (1994) Linkage disequilibrium in growing and stable populations. *Genetics* **137**, 331-336.
Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13486-13491.
Sosef MSM (1994) Studies in Begoniaceae. *Wageningen Ag. University Papers, Netherlands*.
Soto-Calderon ID, Ntie S, Mickala P*, et al.* (2009) Effects of storage type and time on DNA amplification success in tropical ungulate faeces. *Molecular Ecology Resources* **9**, 471-479.
Steffen P, Eggen A, Dietz AB*, et al.* (1993) ISOLATION AND MAPPING OF POLYMORPHIC MICROSATELLITES IN CATTLE. *Animal Genetics* **24**, 121-124.
Swofford D (2000) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland.
Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Symondson WO (2002a) Molecular identification of prey in predator diets. In: *Mol Ecol*, pp. 627-641, England.
Symondson WOC (2002b) Molecular identification of prey in predator diets. *Molecular Ecology* **11**, 627-641.
Taberlet P, Fumagalli L (1996) Owl pellets as a source of DNA for genetic studies of small mammals. *Molecular Ecology* **5**, 301-305.
Taberlet P, Griffin S, Goossens B*, et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**, 3189-3194.
Taberlet P, Luikart G (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society* **68**, 41-55.
Tamura K, Nei M (1993) ESTIMATION OF THE NUMBER OF NUCLEOTIDE SUBSTITUTIONS IN THE CONTROL REGION OF MITOCHONDRIAL-DNA IN HUMANS AND CHIMPANZEES. *Molecular Biology and Evolution* **10**, 512-526.
Tavaré S (1986) Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. In: *American Mathematical Society: Lectures on Mathematics in the Life Sciences*, pp. 57-86. Amer Mathematical Society.
Tchouto MGP, de Wilde JJFE, de Boer WF, van der Maesen LJG, Cleef AM (2009) Bio-indicator species and Central African rain forest refuges in the Campo-Ma'an area, Cameroon. *Systematics and Biodiversity* **7**, 21-31.
Telfer PT, Souquiere S, Clifford SL*, et al.* (2003) Molecular evidence for deep phylogenetic divergence in Mandrillus sphinx. *Molecular Ecology* **12**, 2019-2024.
Thibault M, Blaney S (2003) The oil industry as an underlying factor in the bushmeat crisis in Central Africa. *Conservation Biology* **17**, 1807-1813.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876-4882.
Tomlinson JA, Boonham N, Hughes KJ, Griffin RL, Barker I (2005) On-site DNA extraction and real-time PCR for detection of Phytophthora ramorum in the field. In: *Appl Environ Microbiol*, pp. 6702-6710, United States.

Traub RJ, Robertson ID, Irwin P, Mencke N, Thompson RC (2004a) Application of a species-specific PCR-RFLP to identify Ancylostoma eggs directly from canine faeces. In: *Vet Parasitol*, pp. 245-255, Netherlands.

Traub RJ, Robertson ID, Irwin P, Mencke N, Thompson RCA (2004b) Application of a species-specific PCR-RFLP to identify Ancylostoma eggs directly from canine faeces. *Veterinary Parasitology* **123**, 245.

Travis SE, Keim P (1995) DIFFERENTIATING INDIVIDUALS AND POPULATIONS OF MULE DEER USING DNA. *Journal of Wildlife Management* **59**, 824-831.

Treves A, Mwima P, Plumptre AJ, Isoke S (2010) Camera-trapping forest-woodland wildlife of western Uganda reveals how gregariousness biases estimates of relative abundance and distribution. *Biological Conservation* **143**, 521-528.

Tutin CEG, White LJT, MackangaMissandzou A (1997) The use by rain forest mammals of natural forest fragments in an equatorial African savanna. *Conservation Biology* **11**, 1190-1203.

Umer M, Lamb HF, Bonnefille R*, et al.* (2007) Late pleistocene and holocene vegetation history of the bale mountains, Ethiopia. *Quaternary Science Reviews* **26**, 2229-2246.

Vaiman D, Mercier D, Moazami-Goudarzi K*, et al.* (1994) A set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism. *Mamm Genome* **5**, 288-297.

Valiere N (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes* **2**, 377-379.

Valiere N, Fumagalli L, Gielly L*, et al.* (2003) Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. *Animal Conservation* **6**, 83-92.

Vallet D, Petit EJ, Gatti S, Levrero F, Menard N (2008) A new 2CTAB/PCI method improves DNA amplification success from faeces of Mediterranean (Barbary macaques) and tropical (lowland gorillas) primates. *Conservation Genetics* **9**, 677-680.

Van Hooft WF, Groen AF, Prins HT (2002) Phylogeography of the African buffalo based on mitochondrial and Y-chromosomal loci: Pleistocene origin and population expansion of the Cape buffalo subspecies. *Molecular Ecology* **11**, 267-279.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

van Vliet N, Nasi R (2008a) Hunting for Livelihood in Northeast Gabon: Patterns, Evolution, and Sustainability. *Ecology and Society* **13**, -.

van Vliet N, Nasi R (2008b) Why do models fail to assess properly the sustainability of duiker (Cephalophus spp.) hunting in Central Africa? *Oryx* **42**, 392-399.

van Vliet N, Zundel S, Miquel C, Taberlet P, Nasi R (2008) Distinguishing dung from blue, red and yellow-backed duikers through noninvasive genetic techniques. *African Journal of Ecology* **46**, 411-417.

van Vuuren BJ, Robinson TJ (2001) Retrieval of four adaptive lineages in duiker antelope: Evidence from mitochondrial DNA sequences and fluorescence in situ hybridization. *Molecular Phylogenetics and Evolution* **20**, 409-425.

Vanden Bossche J-P, Bernacsek GM (1990) Source book for the inland fishery resources of Africa: 2. In: *CIFA Technical Paper*, p. 411. FAO, Rome.

Vazquez-Dominguez E, Paetkau D, 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). *Molecular Ecology* **10**, 2069-2078.

Vincens A (1991) Late Quaternary Vegetation History of the South-Tanganyika Basin - Climatic Implications in South-Central-Africa. *Palaeogeography Palaeoclimatology Palaeoecology* **86**, 207-226.
Vincens A, Buchet G, Williamson D, Taieb M (2005) A 23,000 yr pollen record from Lake Rukwa (8 degrees S, SW Tanzania): New data on vegetation dynamics and climate in Central Eastern Africa. *Review of Palaeobotany and Palynology* **137**, 147-162.
Vincens A, Garcin Y, Buchet G (2007) Influence of rainfall seasonality on African lowland vegetation during the Late Quaternary: pollen evidence from Lake Masoko, Tanzania. *Journal of Biogeography* **34**, 1274-1288.
Viteri MP, Waits LP (2009) Identifying polymorphic microsatellite loci for Andean bear research. *Ursus* **20**, 102-108.
Vrba ES (1995) The fossil record of African antelopes (Mammalia, Bovidae) in relation to human evolution and paleoclimate. In: *Paleoclimate and Evolution with emphasis on Human Origins* (ed. E. S. Vrba GHD, T. C. Partridge and L. H. Burckle, L.H.), pp. 385-424. Yale University Press, New Haven, Connecticut.
Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. In: *Mol Ecol*, pp. 249-256, England.
Waits LP, Paetkau D (2005a) Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *Journal of Wildlife Management* **69**, 1419.
Waits LP, Paetkau D (2005b) Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *Journal of Wildlife Management* **69**, 1419-1433.
Walker KL (2006) Impact of the little fire ant, Wasmannia auropunctata, on native forest ants in Gabon. *Biotropica* **38**, 666-673.
Wallace AR (1852) On the monkeys of the Amazon. *Proceedings of the Zoological Society of London* **20**, 107-110.
Walsh PD, White LJT (1999) What it will take to monitor forest elephant populations. *Conservation Biology* **13**, 1194-1202.
Waltert M, Heber S, Riedelbauch S, Lien JL, Muhlenberg M (2006) Estimates of blue duiker (Cephalophus monticola) densities from diurnal and nocturnal line transects in the Korup region, south-western Cameroon. *African Journal of Ecology* **44**, 290-292.
Wasser SK, Houston CS, Koehler GM, Cadd GG, Fain SR (1997) Techniques for application of faecal DNA methods to field studies of Ursids. *Molecular Ecology* **6**, 1091-1097.
Wasser SK, Mailand C, Booth R*, et al.* (2007) Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 4228-4233.
Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical population biology* **7**, 256-276.
White L, Edwards, A. (2000) *Conservation research in the African rain forests a technical handbook* The Wildlife Conservation Society New York, U. S. A.
Wiens JJ, Penkrot TA (2002) Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). *Systematic Biology* **51**, 69-91.
Wilkie DS, Carpenter JF (1999) Bushmeat hunting in the Congo Basin: an assessment of impacts and options for mitigation. *Biodiversity and Conservation* **8**, 927-955.
Wilkie DS, Curran B, Tshombe R, Morelli GA (1998) Managing bushmeat hunting in Okapi Wildlife Reserve, Democratic Republic of Congo. *Oryx* **32**, 131-144.
Wilkie DS, Sidle JG, Boundzanga GC (1992) MECHANIZED LOGGING, MARKET HUNTING, AND A BANK LOAN IN CONGO. *Conservation Biology* **6**, 570-580.

Wilkie DS, Starkey M, Abernethy K*, et al.* (2005) Role of prices and wealth in consumer demand for bushmeat in Gabon, central Africa. *Conservation Biology* **19**, 268-274.
Willows-Munro S, Robinson TJ, Matthee CA (2005) Utility of nuclear DNA intron markers at lower taxonomic levels: Phylogenetic resolution among nine Tragelaphus spp. *Molecular Phylogenetics and Evolution* **35**, 624-636.
Wronski T, Hausdorf B (2008) Distribution patterns of land snails in Ugandan rain forests support the existence of Pleistocene forest refugia. *Journal of Biogeography* **35**, 1759-1768.
Zapata MA, Cienfuegos AV, Quiros OI*, et al.* (2007) Discrimination of seven Anopheles species from San Pedro De Uraba, Antioquia, Colombia, by polymerase chain reaction-restriction fragment length polymorphism analysis of its sequences. *American Journal of Tropical Medicine and Hygiene* **77**, 67-72.

Appendices

Chapter 2 appendices

Table S1. McNemar test: pair-wise *post hoc* comparisons of differences in mitochondrial amplification success between treatments (storage type and time).

Time	Exact Sig. (2-tailed)	Storage method	Exact Sig. (2-tailed)
SG1 & SG2	**<0.001**	SG1 & RNA1	0.625
SG1 & SG3	**<0.001**	SG1 & ET1	0.625
SG2 & SG3	0.146	RL1 & ET1	1.000
RL1 & RL2	0.250	SG2 & RL2	**<0.001**
RL1 & RL3	**<0.001**	SG2 & ET2	0.388
RL2 & RL3	**0.001**	RL2 & ET2	**<0.001**
ET1 & ET2	**<0.001**	SG3 & RL3	**<0.001**
ET1 & ET3	**<0.001**	SG3 & ET3	0.453
ET2 & ET3	0.065	RL3 & ET3	**0.012**

1 = one week, 2 = one month, 3 = 3 months, SG = Silica, RL = RNA*later*, ET = ethanol. Significant values after Bonferroni-Holm correction are in **bold**.

Table S2. Wilcoxon Test *post hoc* comparisons of differences in nuclear DNA concentrations between treatments (storage type and time).

Time	Z	Asymptotic Sig. (2-tailed)	Storage type	Z	Asymptotic Sig. (2-tailed)
SG1 & SG2	-0.508	0.611	RNA1 & SG1	-4.357	**<0.001**
SG1 & SG3	-1.373	0.170	ETOH1 & SG1	-4.304	**<0.001**
SG2 & SG3	-0.811	0.417	ETOH1 & RNA1	-0.032	0.974
RNA1 & RNA2	-0.843	0.400	RNA2 & SG2	-4.236	**<0.001**
RNA1 & RNA3	-0.192	0.848	ETOH2 & SG2	-4.531	**<0.001**
RNA2 & RNA3	-0.865	0.387	ETOH2 & RNA2	-3.427	**<0.001**
ETOH1 & ETOH2	-3.125	**0.002**	RNA3 & SG3	-4.379	**<0.001**
ETOH1 & ETOH3	-3.254	**0.001**	ETOH3 & SG3	-4.552	**<0.001**
ETOH2 & ETOH3	-0.934	0.350	ETOH3 & RNA3	-3.051	**0.002**

1 = one week, 2 = one month, 3 = 3 months, SG = Silica, RNA = RNAlater, ETOH = ethanol. Significant values after Bonferroni-Holm correction are in **bold**.

Table S3. Wilcoxon pair-wise *post hoc* comparisons of differences in nuclear microsatellite amplification rates between treatments (storage type and time).

Time	Z	Asymptotic Sig. (2-tailed)	Storage type	Z	Asymptotic Sig. (2-tailed)
SG1 & SG2	-0.707	0.479	RNA1 & SG1	-3.194	**0.001**
SG1 & SG3	-0.301	0.763	ETOH1 & SG1	-2.936	**0.003**
SG2 & SG3	-0.333	0.739	RNA2 & SG2	-2.874	**0.004**
RNA1 & RNA2	-1.231	0.218	ETOH2 & SG2	-3.762	**<0.001**
RNA1 & RNA3	-1.545	0.122	RNA3 & SG3	-2.164	0.030
RNA2 & RNA3	-0.209	0.834	ETOH3 & SG3	-3.878	**<0.001**
ETOH1 & ETOH2	-1.784	0.074	ETOH1 & RNA1	-0.554	0.579
ETOH1 & ETOH3	-2.888	**0.004**	ETOH2 & RNA2	-2.519	**0.012**
ETOH2 & ETOH3	-2.369	**0.018**	ETOH3 & RNA3	-4.000	**<0.001**

1 = one week, 2 = one month, 3 = 3 months, SG = Silica, RNA-L = RNALater, ETOH = ethanol. Significant values are in bold.

Table S4. Wilcoxon pair-wise *post hoc* comparisons of differences in recovery of consensus genotypes in nuclear microsatellites between treatments (storage type and time).

Time	Z	Asymptotic Sig. (2-tailed)	Storage type	Z	Asymptotic Sig. (2-tailed)
SG1 & SG2	-1.147	0.251	SG1 & RNA1	-3.916	**<0.001**
SG1 & SG3	-2.154	0.031	SG1 & ETOH1	-3.476	**<0.001**
SG2 & SG3	-.954	0.340	RNA1 & ETOH1	-1.055	0.292
RNA1 & RNA2	-2.361	0.018	SG2 & RNA2	-3.032	**0.002**
RNA1 & RNA3	-.545	0.586	SG2 & ETOH2	-4.470	**<0.001**
RNA2 & RNA3	-2.143	0.032	RNA2 & ETOH2	-3.845	**<0.001**
ETOH1 & ETOH2	-2.965	**0.003**	SG3 & RNA3	-3.487	**<0.001**
ETOH1 & ETOH3	-3.583	**<0.001**	SG3 & ETOH3	-4.525	**<0.001**
ETOH2 & ETOH3	-2.172	0.030	RNA3 & ETOH3	-3.463	**<0.001**

1 = one week, 2 = one month, 3 = 3 months, SG = Silica, RNA-L = RNALater, ETOH = ethanol. Significant values after Bonferroni-Holm correction are in **bold**.

Table S5. McNemar *post hoc* comparisons of differences in individual microsatellite amplification success between time periods.

Within locus comparisons	Exact Sig. (2-tailed)		
	Silica	RNA*later*	ETOH
INRA40_1 & INRA40_2	1.000	0.219	0.625
INRA40_1 & INRA40_3	1.000	0.125	0.021
INRA40_2 & INRA40_3	1.000	1.000	0.109
BM1225_1 & BM1225_2	0.500	0.219	0.146
BM1225_1 & BM1225_3	1.000	0.219	0.022
BM1225_2 & BM1225_3	1.000	1.000	0.607
BM2113_1 & BM2113_2	1.000	1.000	0.125
BM2113_1 & BM2113_3	1.000	1.000	0.021
BM2113_2 & BM2113_3	1.000	0.625	0.453
BRIBBO_1 & BRIBBO_2	0.250	0.688	0.549
BRIBBO_1 & BRIBBO_3	1.000	0.508	0.049
BRIBBO_2 & BRIBBO_3	0.500	1.000	0.146
BM1862_1 & BM1862_2	1.000	0.289	0.688
BM1862_1 & BM1862_3	1.000	0.727	0.118
BM1862_2 & BM1862_3	1.000	0.727	0.004
BM143_1 & BM143_2	0.250	1.000	1.000
BM143_1 & BM143_3	1.000	1.000	0.791
BM143_2 & BM143_3	0.688	1.000	0.754

1 = one week, 2 = one month, 3 = 3 months. Significant values after Bonferroni-Holm correction are in **bold**.

Chapter 3 appendices

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.

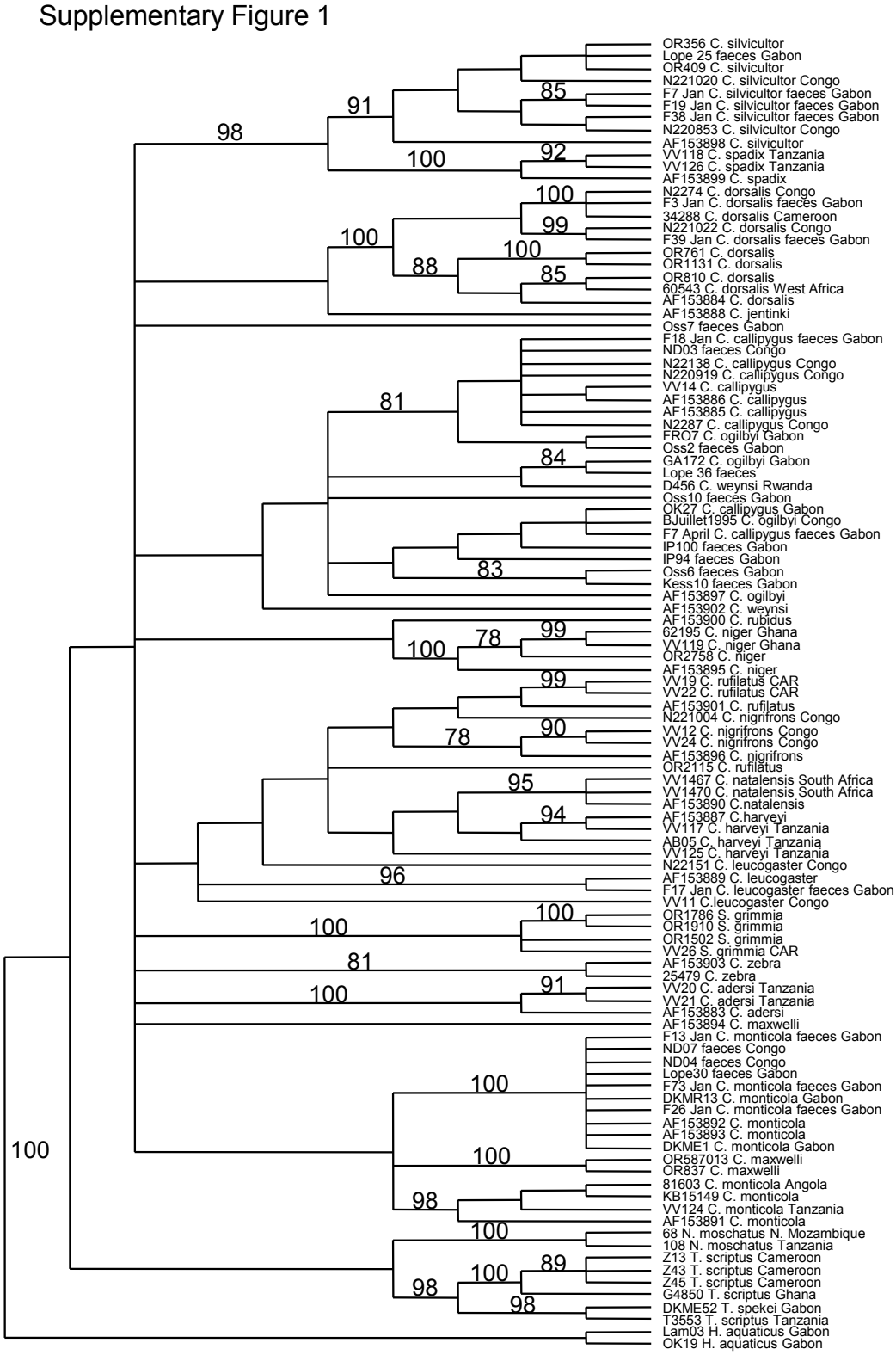


Figure S2. Cytochrome b maximum parsimony bootstrap consensus phylogeny rooted with H. aquaticus. Sequences were aligned with CLUSTAL. Bootstrap values are indicated at the appropriate node.

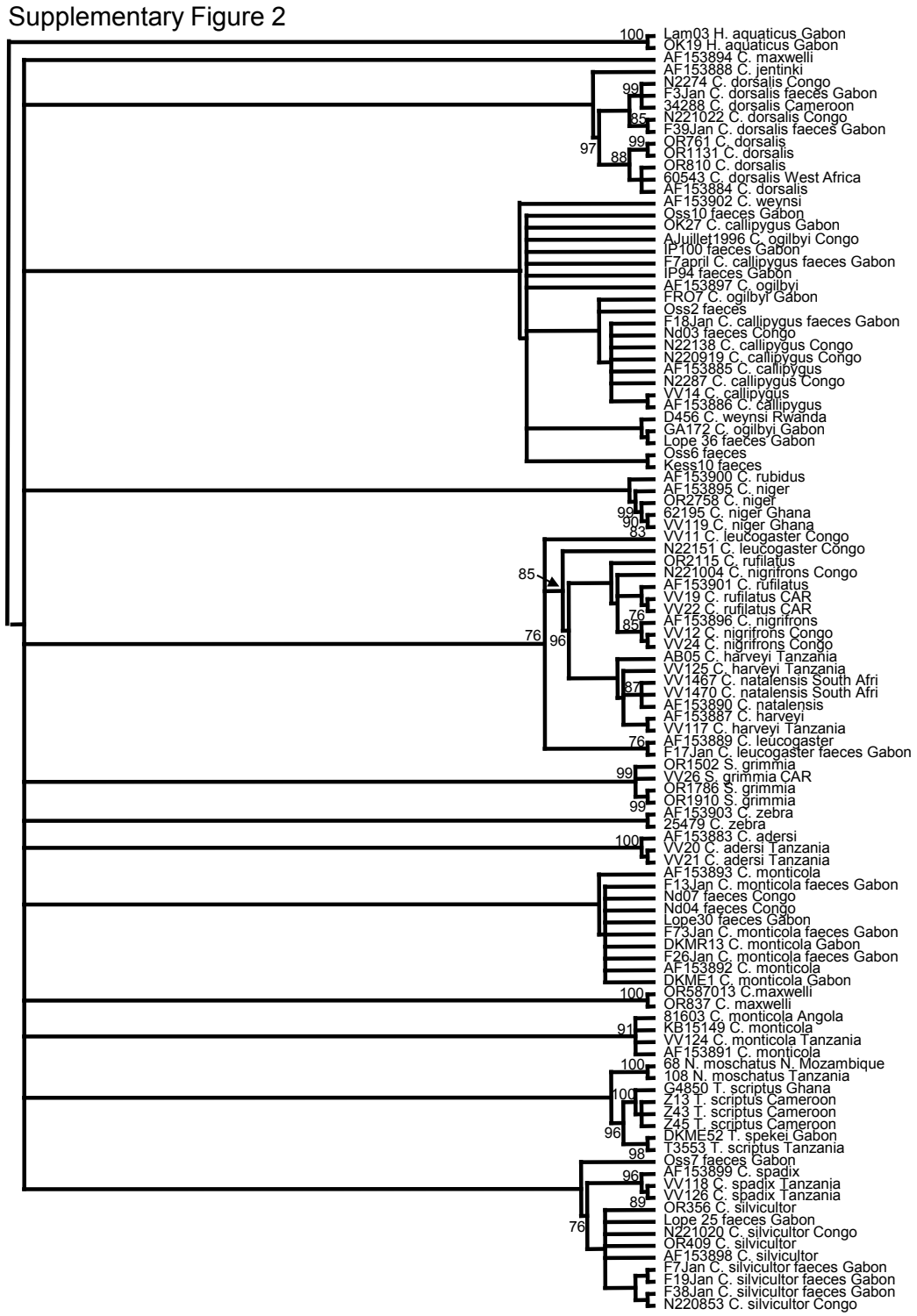


Figure S3. Control region phylogeny based on Bayesian analysis and rooted with *Tragelaphus* species. Sequences were aligned with CLUSTAL. Posterior probability values are indicated at the appropriate node.

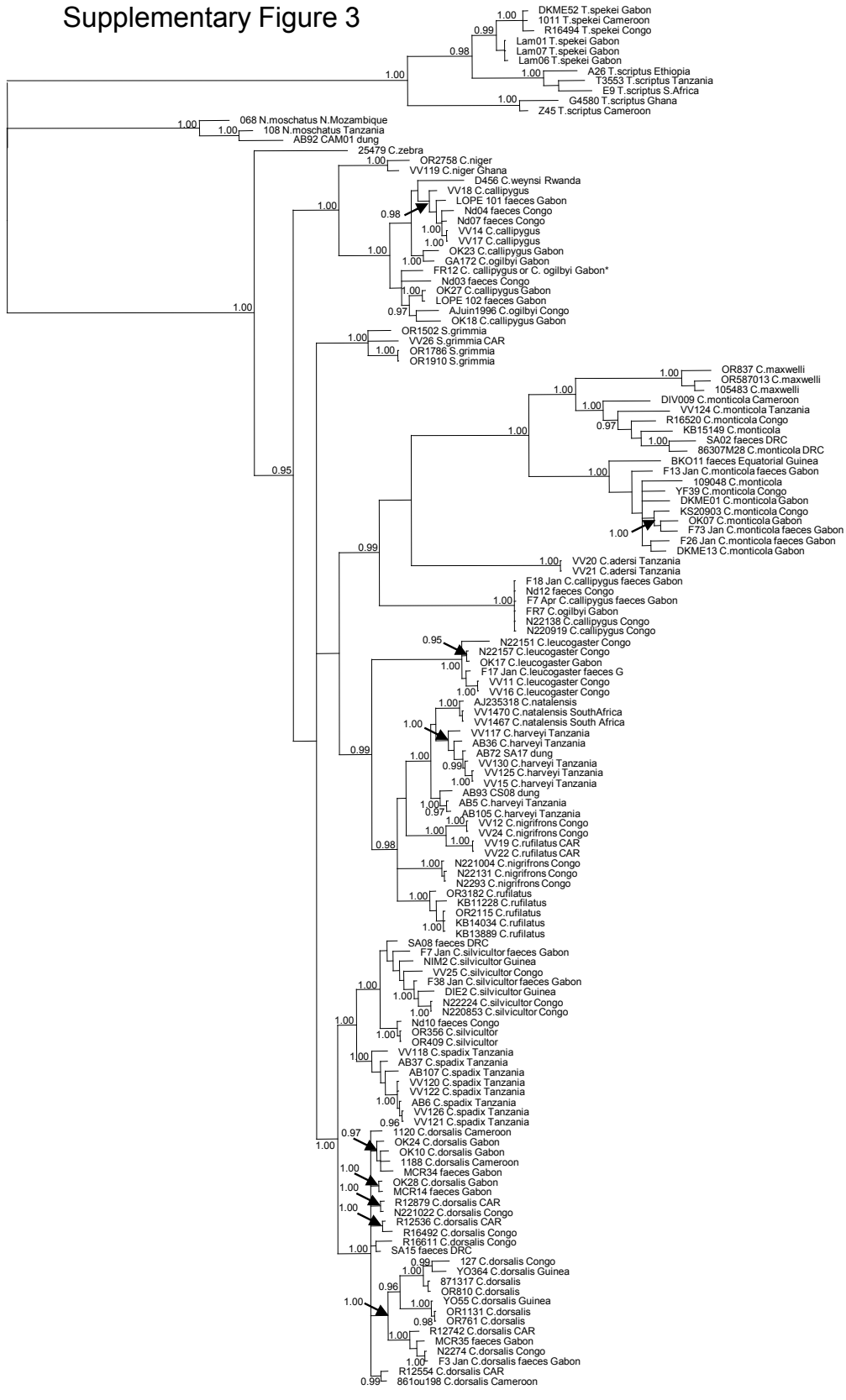


Figure S4. Control region neighbour-joining bootstrap consensus phylogeny based on Kimura-2-parameter corrected distances and rooted with *N. moschatus*. Sequences were aligned with CLUSTAL. Bootstrap values are indicated at the relevant node.

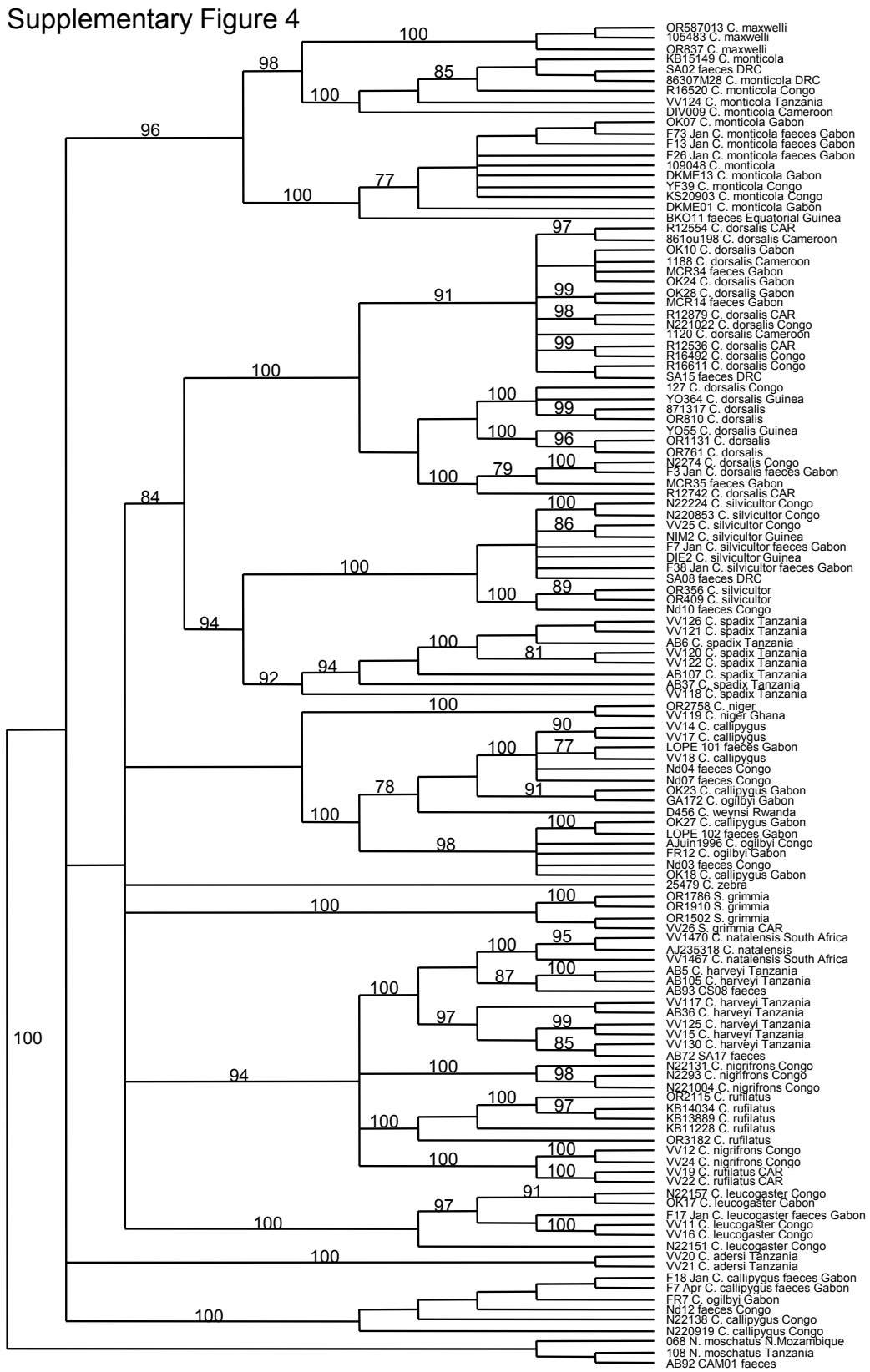


Figure S5. Control region maximum parsimony bootstrap consensus phylogeny rooted with *N. moschatus*. Sequences were aligned with CLUSTAL. Bootstrap values are indicated at the appropriate node.

Supplementary Figure 5

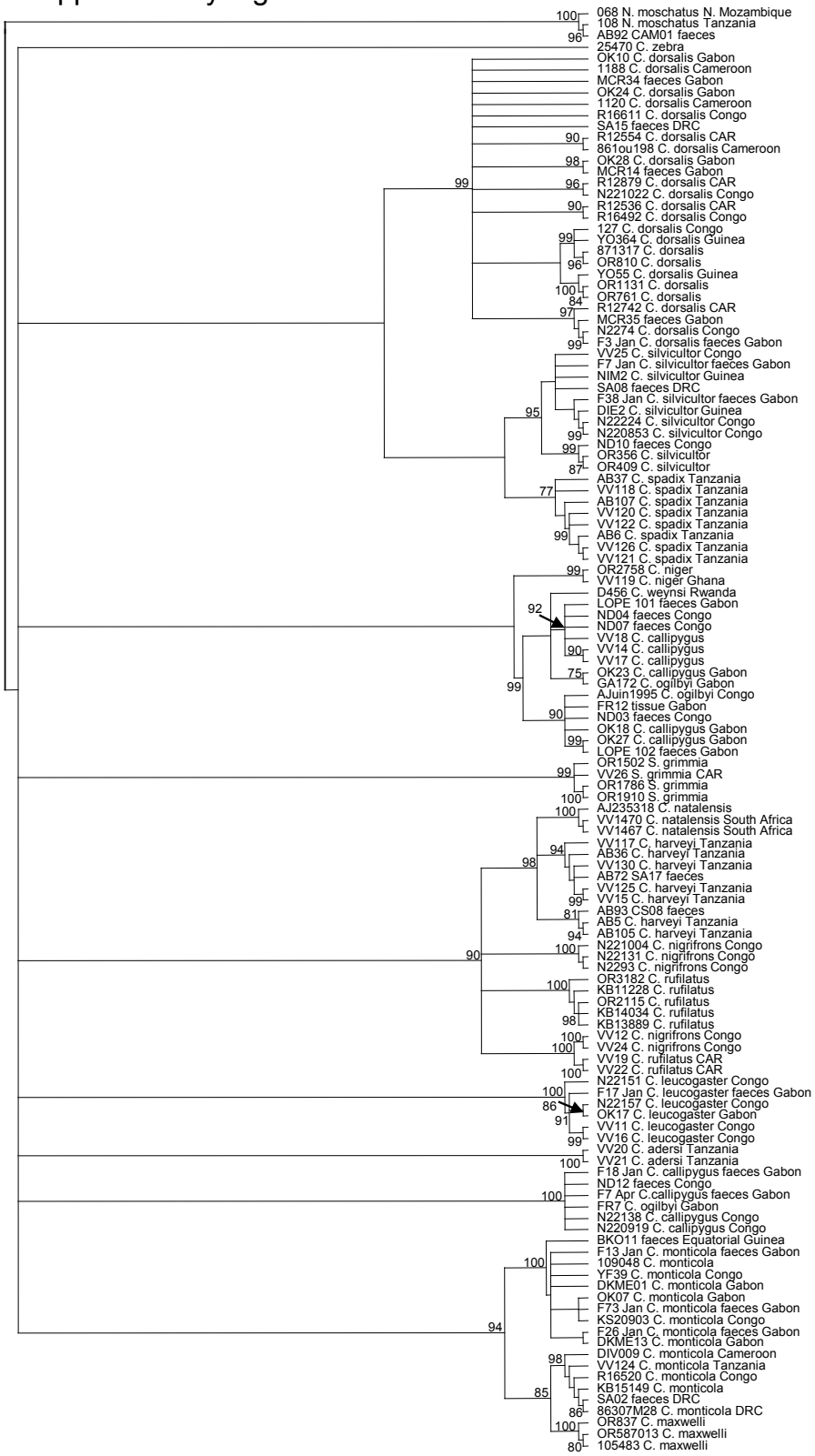


Figure S6. Flow chart illustrating the application of RFLP banding patterns to the diagnosis of species using cytochrome b gene sequence data.

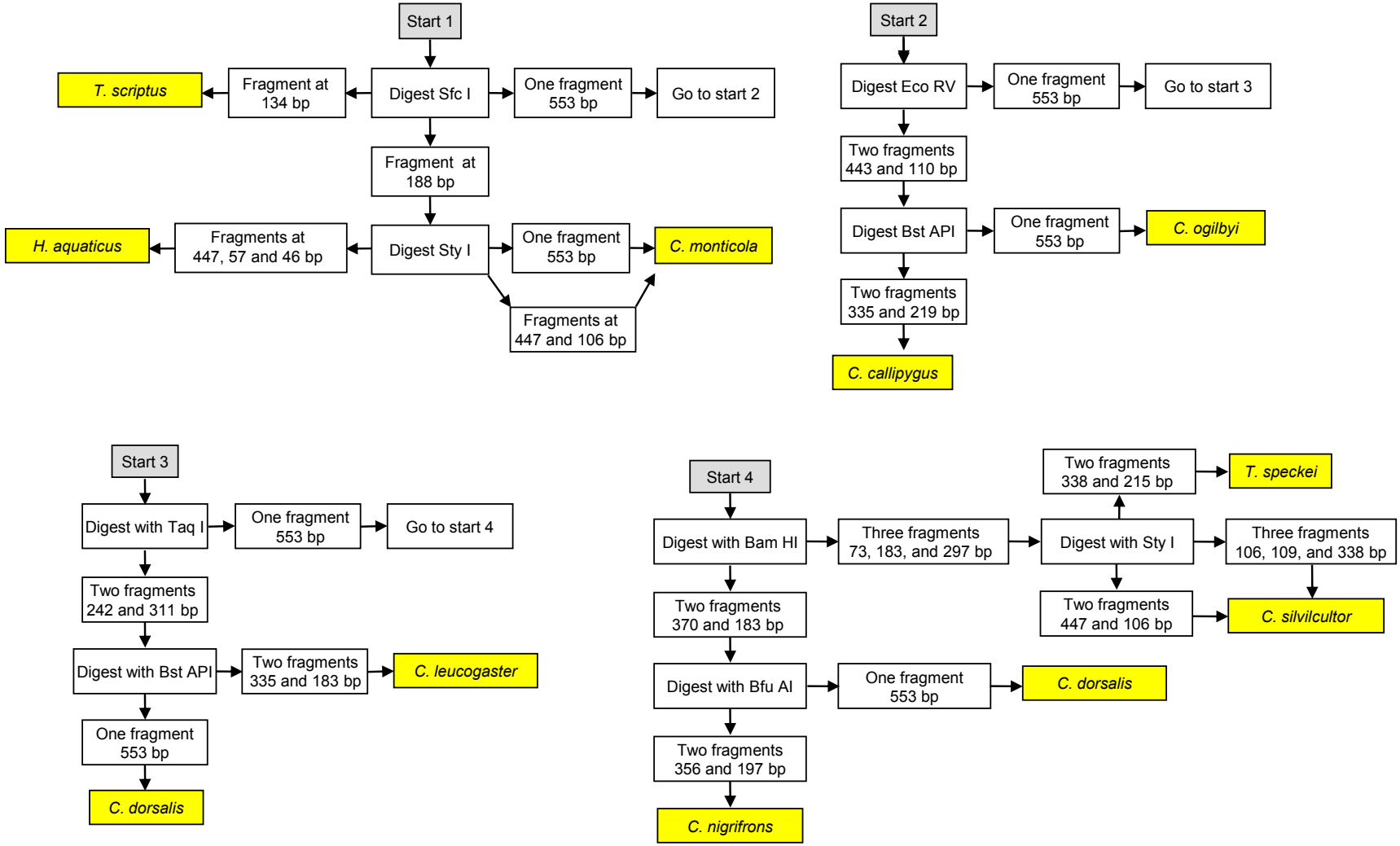


Table S1. Details of the species, provider, geographic origin and GenBank accession numbers for samples used in the present study.

Sample code	Provider	Species	Geographic origin	Control region	Cyt b
Z45	Y. Moodley	*T. scriptus*	Bipindi, Cameroon	FJ823276	FJ807582
A26	Y. Moodley	*T. scriptus*	Cure-Rey, Ethiopia	FJ823277	
Z43	Y. Moodley	*T. scriptus*	Kribi, Cameroon		FJ807581
Z13	Y. Moodley	*T. scriptus*	Bipindi, Cameroon		FJ807580
T3553	Y. Moodley	*T. scriptus*	Ugalla West, Tanzania	FJ823278	FJ807604
G4580	Y. Moodley	*T. scriptus*	Ashanti Region, Ghana	FJ823279	FJ807583
E9	Y. Moodley	*T.scriptus*	Eastern Cape, S. Africa	FJ82380	
R16494	M. Colyn	*T. spekei*	Lefini, Republic of the Congo	FJ823281	
1011	M. Colyn	*T. spekei*	Dja, Cameroon	FJ823282	
DKME52	S. Touladjan	*T. spekei*	Okondja, Gabon	FJ823283	FJ807594
Lam06	S. Touladjan	*T. spekei*	Lambaréné, Gabon	FJ823284	
Lam07	S. Touladjan	*T. spekei*	Lambaréné, Gabon	FJ823285	
Lam01	S. Touladjan	*T. spekei*	Lambaréné, Gabon	FJ823286	
Lam03	S. Touladjan	*H. aquaticus*	Lambaréné, Gabon	FJ823287	FJ807599
FR19	S. Touladjan	*H. aquaticus*	Franceville, Gabon	FJ823288	
Lam04	S. Touladjan	*H. aquaticus*	Lambaréné, Gabon	FJ823289	
OK15	S. Touladjan	*H. aquaticus*	Okondja, Gabon	FJ823290	
OK19	S. Touladjan	*H. aquaticus*	Okondja, Gabon	FJ823291	FJ807600
FR05	S. Touladjan	*H. aquaticus*	Franceville, Gabon	FJ823292	
FR16	S. Touladjan	*H. aquaticus*	Franceville, Gabon	X	
457	M. Colyn	*H. aquaticus*	Dja, Cameroon	FJ823293	
OR1786	San Diego Zoo	*S. grimmia*		FJ823294	FJ807591
OR1910	San Diego Zoo	*S. grimmia*	S. Africa	FJ823295	FJ807592
OR1502	San Diego Zoo	*S. grimmia*		FJ823296	FJ807593
VV26	B.J. van Vuuren	*S. grimmia*	CAR	FJ823297	FJ807613
068	A. Bowkett	*N. moschatus*	Mozambique	FJ985772	FJ959386
108	A. Bowkett	*N. moschatus*	Tanzania	FJ985773	FJ959387
YF39	D. Pires	*C. monticola*	Republic of the Congo	FJ823298	
OK07	S. Touladjan	*C. monticola*	Okondja, Gabon	FJ823299	
KS20903	D. Pires	*C. monticola*	Republic of the Congo	FJ823300	
VV124	B.J. van Vuuren	*C. monticola*	Tanzania	FJ823301	FJ807607
KB15149	San Diego Zoo	*C. monticola*	Cape Province, S. Africa	FJ823302	FJ807619
109048	AMNH	*C. monticola*		FJ823303	
DKMR13	S. Touladjan	*C. monticola*	Gabon	FJ823304	FJ807584
DKME01	S. Touladjan	*C. monticola*	Gabon	FJ823305	FJ807585
86307M28	M. Colyn	*C. monticola*	Kinsangani, DRC	FJ823306	
DIV009	M. Colyn	*C. monticola*	Bamenda, Cameroon	FJ823307	
R16520	M. Colyn	*C. monticola*	Lefini, Republic of the Congo	FJ823308	
81603	Field Museum	*C. monticola*	Malanje, Angola,		FJ807598
AF153893	Genbank	*C. monticola*	Cameroon		AF153893

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

AF153892	Genbank	*C. monticola*	Gabon		AF153892
AF153891	Genbank	*C. monticola*	S. Africa		AF153891
AF153894	Genbank	*C. maxwelli*	Ghana		AF153894
105483	AMNH	*C. maxwelli*		FJ823309	
OR587013	San Diego Zoo	*C. maxwelli*		FJ823310	FJ807602
OR837	San Diego Zoo	*C. maxwelli*		FJ823311	FJ807603
VV21	B.J. van Vuuren	*C. adersi*	Zanzibar, Tanzania	FJ823312	FJ807617
VV20	B.J. van Vuuren	*C. adersi*	Zanzibar, Tanzania	FJ823313	FJ807616
AF153883	Genbank	*C. adersi*	Tanzania	AF153883	AF153883
VV1470	B.J. van Vuuren	*C. natalensis*	KwaZulu-Natal, S. Africa	FJ823314	FJ807611
VV1467	B.J. van Vuuren	*C. natalensis*	KwaZulu-Natal, S. Africa	FJ823315	FJ807610
AJ235318	GenBank	*C. natalensis*		AJ235318	
AF153890	Genbank	*C. natalensis*	S. Africa		AF153890
140902	Field Museum	*C. natalensis*		X	
VV125	B.J. van Vuuren	*C. harveyi*	Mt. Meru, Tanzania	FJ823316	FJ807615
VV117	B.J. van Vuuren	*C. harveyi*	Usambara Mts., Tanzania	FJ823317	FJ807623
VV15	B.J. van Vuuren	*C. harveyi*	Mt. Meru, Tanzania	FJ823318	
VV130	B.J. van Vuuren	*C. harveyi*	Usambara Mts., Tanzania	FJ823319	
AB5	F. Rovero	*C. harveyi*	Udzungwa Mts., Tanzania	AM903088	FJ959388
AB36	F. Rovero	*C. harveyi*	Rubeho Mts., Tanzania	AM903089	
AB105	A. Bowkett	*C. harveyi*	Udzungwa Mts., Tanzania	AM903090	
AF153887	Genbank	*C. harveyi*	Tanzania		AF153887
OR2115	San Diego Zoo	*C. rufilatus*		FJ823320	FJ807586
VV19	B.J. van Vuuren	*C. rufilatus*	CAR	FJ823321	FJ807625
VV22	B.J. van Vuuren	*C. rufilatus*	CAR	FJ823322	FJ807626
KB11228	San Diego Zoo	*C. rufilatus*	Guinea	FJ823323	
KB13889	San Diego Zoo	*C. rufilatus*		FJ823324	
KB14034	San Diego Zoo	*C. rufilatus*		FJ823325	
OR3182	San Diego Zoo	*C. rufilatus*	Guinea	FJ823326	
AF153901	Genbank	*C. rufilatus*	CAR		AF153901
N221004	D. Pires	*C. nigrifrons*	Republic of the Congo	FJ823327	FJ807572
VV12	B.J. van Vuuren	*C. nigrifrons*	Republic of the Congo	FJ823328	FJ807609
N2293	D. Pires	*C. nigrifrons*	Republic of the Congo	FJ823329	
N22131	D. Pires	*C. nigrifrons*	Republic of the Congo	FJ823330	
VV24	B.J. van Vuuren	*C. nigrifrons*	Republic of the Congo	FJ823331	FJ807627
AF153896	Genbank	*C. nigrifrons*	DRC		AF153896
25479	San Diego Zoo	*C. zebra*	Liberia	FJ823332	FJ807601
AF153903	Genbank	*C. zebra*	Liberia		AF153903
T14	Marc Colyn	*C. zebra*	Côte d'Ivoire	X	
VV11	B.J. van Vuuren	*C. leucogaster*	Republic of the Congo	FJ823333	FJ807606
VV16	B.J. van Vuuren	*C. leucogaster*	Republic of the Congo	FJ823334	

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

N22157	D. Pires	*C. leucogaster*	Republic of the Congo	FJ823335	
N22151	D. Pires	*C. leucogaster*	Republic of the Congo	FJ823336	FJ807578
OK17	S. Touladjan	*C. leucogaster*	Okondja, Gabon	FJ823337	
AF153889	Genbank	*C. leucogaster*	DRC		AF153889
VV14	B.J. van Vuuren	*C. callipygus*	Republic of the Congo	FJ823338	FJ807612
VV17	B.J. van Vuuren	*C. callipygus*	Republic of the Congo	FJ823339	
VV18	B.J. van Vuuren	*C. callipygus*	Republic of the Congo	FJ823340	
OK23	S. Touladjan	*C. callipygus*	Okondja, Gabon	FJ823341	
OK27	S. Touladjan	*C. callipygus*	Okondja, Gabon	FJ823342	FJ807620
N22138	D. Pires	*C. callipygus*	Republic of the Congo	FJ823343	FJ807573
N220919	D. Pires	*C. callipygus*	Republic of the Congo	FJ823344	FJ807574
N2287	D. Pires	*C. callipygus*	Republic of the Congo		FJ807575
OK18	S. Touladjan	*C. callipygus*	Okondja, Gabon	FJ823345	
AF153886	Genbank	*C. callipygus*	DRC		AF153886
AF153885	Genbank	*C. callipygus*	Gabon		AF153885
OR2758	San Diego Zoo	*C. niger*	Liberia	FJ823346	FJ807621
VV119	B.J. van Vuuren	*C. niger*	Ghana	FJ823347	FJ807624
VV131	B.J. van Vuuren	*C. niger*	Ghana	X	
AF153895	Genbank	*C. niger*	Ghana		AF153895
62195	Field Museum	*C. niger*	Ghana		FJ807597
VV118	B.J. van Vuuren	*C. spadix*	Kilimanjaro West, Tanzani	FJ823348	FJ807605
VV122	B.J. van Vuuren	*C. spadix*	Usambara Mts., Tanzania	FJ823349	
VV121	B.J. van Vuuren	*C. spadix*	Usambara Mts., Tanzania	FJ823350	
VV120	B.J. van Vuuren	*C. spadix*	Usambara Mts., Tanzania	FJ823351	
AB6	J. Beraducci	*C. spadix*	W. Usambaras, Tanzania	AM903084	
AB37	T. Davenport	*C. spadix*	S. highlands, Tanzania	AM903085	
AB107	T. Davenport	*C. spadix*	S. highlands, Tanzania	AM903086	
AF153899	Genbank	*C. spadix*	Tanzania		AF153899
VV126	B.J. van Vuuren	*C. spadix*	Usambara Mts., Tanzania	FJ823352	FJ807608
N22224	D. Pires	*C. silvicultor*	Republic of the Congo	FJ823353	
OR356	San Diego Zoo	*C. silvicultor*	Liberia	FJ823354	FJ807622
VV25	B.J. van Vuuren	*C. silvicultor*	Republic of the Congo	FJ823355	
OR409	San Diego Zoo	*C. silvicultor*	Liberia	FJ823356	FJ807587
N220853	D. Pires	*C. silvicultor*	Republic of the Congo	FJ823357	FJ807579
DIE2	M. Colyn	*C. silvicultor*	Diecke, Guinea	FJ823358	
NIM2	M. Colyn	*C. silvicultor*	Mt. Nimba, Guinea	FJ823359	
N221020	D. Pires	*C. silvicultor*	Republic of the Congo		FJ807571
AF153898	Genbank	*C. silvicultor*			AF153898
AJuin1995	M. Colyn	*C. ogilbyi*	Brazzaville, Republic of the Congo	FJ823360	
BJuillet1995	M. Colyn	*C. ogilbyi*	Odzala, Republic of the Congo		FJ888512
FR7	S. Touladjan	*C. ogilbyi*	Franceville, Gabon	FJ823361	FJ807618

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

FR12	S. Touladjan	*C. ogilbyi or C. callipygus*	Franceville, Gabon	FJ823362	
GA172	M. Colyn	*C. ogilbyi*	Maloungа, Gabon	FJ823363	FJ807628
AF153897	Genbank	*C. ogilbyi*	Nigeria		AF153897
OK28	S. Touladjan	*C. dorsalis*	Okondja, Gabon	FJ823364	
N221022	D. Pires	*C. dorsalis*	Republic of the Congo	FJ823365	FJ807577
N2274	D. Pires	*C. dorsalis*	Republic of the Congo	FJ823366	FJ807576
OR761	San Diego Zoo	*C. dorsalis*		FJ823367	FJ807588
OR1131	San Diego Zoo	*C. dorsalis*		FJ823368	FJ807589
OR810	San Diego Zoo	*C. dorsalis*		FJ823369	FJ807590
871317	AMNH	*C. dorsalis*		FJ823370	
34288	Field Museum	*C. dorsalis*	Cameroon		FJ807595
60543	Field Museum	*C. dorsalis*	West Africa		FJ807596
R16611	M. Colyn	*C. dorsalis*	Lefini, Republic of the Congo	FJ823371	
YO55	M. Colyn	*C. dorsalis*	Diecke, Guinea	FJ823372	
R16492	M. Colyn	*C. dorsalis*	Lefini, Republic of the Congo	FJ823373	
R12742	M. Colyn	*C. dorsalis*	Ngotto, CAR	FJ823374	
1120	M. Colyn	*C. dorsalis*	Dja, Cameroon	FJ823375	
861ou198	M. Colyn	*C. dorsalis*	Dja, Cameroon	FJ823376	
YO364	M. Colyn	*C. dorsalis*	Diecke, Guinea	FJ823377	
OK10	S. Touladjan	*C. dorsalis*	Okondja, Gabon	FJ823378	
OK24	S. Touladjan	*C. dorsalis*	Okondja, Gabon	FJ823379	
127	M. Colyn	*C. dorsalis*	Odzala, Republic of the Congo	FJ823380	
R12536	M. Colyn	*C. dorsalis*	Ngotto, CAR	FJ823381	
R12554	M. Colyn	*C. dorsalis*	Ngotto, CAR	FJ823382	
R12879	M. Colyn	*C. dorsalis*	Ngotto, CAR	FJ823383	
1188	M. Colyn	*C. dorsalis*	Dja, Cameroon	FJ823384	
AF153884	Genbank	*C. dorsalis*	Ghana		AF153884
0108-1DOR	M. Colyn	*C. dorsalis*	Mt. Nimba, Guinea	X	
D456	B.J. van Vuuren	*C. weynsi*	Rwanda	FJ823385	FJ807614
AF153902	Genbank	*C. weynsi*	Rwanda		AF153902
AF153888	Genbank	*C. jentinki*			AF153888
AF153900	Genbank	*C. rubidus*	Uganda		AF153900
F3Jan	N. van Vliet *et al.* 2008a	Putatively *C. dorsalis*	Ivindo, Gabon	X	X
F39Jan	N. van Vliet *et al.* 2008a	Putatively *C. dorsalis*	Ivindo, Gabon		X
F26Jan	N. van Vliet *et al.* 2008a	Putatively *C. monticola*	Ivindo, Gabon	X	X
F13Jan	N. van Vliet *et al.* 2008a	Putatively *C. monticola*	Ivindo, Gabon	X	X
F73Jan	N. van Vliet *et al.* 2008a	Putatively *C. monticola*	Ivindo, Gabon	X	X
F7April	N. van Vliet *et al.* 2008a	Putatively *C. callipygus*	Ivindo, Gabon	X	X
F18Jan	N. van Vliet *et al.* 2008a	Putatively *C. callipygus*	Ivindo, Gabon	X	X
F17Jan	N. van Vliet *et al.* 2008a	Putatively *C. leucogaster*	Ivindo, Gabon	X	X
F7Jan	N. van Vliet *et al.* 2008a	Putatively *C. silvicultor*	Ivindo, Gabon	X	X

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

F38Jan	N. van Vliet *et al.* 2008a	Putatively *C. silvicultor*	Ivindo, Gabon	X	X
F19Jan	N. van Vliet *et al.* 2008a	Putatively *C. silvicultor*	Ivindo, Gabon		X
BKO11	A. Johnston, S. Ntie	Faeces of unknown ID	Bioko Island, Eq. Guinea	X	
SA02	B. Curran	Faeces of unknown ID	Salonga, DRC	X	
AB72_SA17	A. Bowkett	Faeces of unknown ID	Udzungwa Mts., Tanzania	X	
AB92_CAM01	A. Bowkett	Faeces of unknown ID	Udzungwa Mts., Tanzania	X	
AB93_CS08	A. Bowkett	Faeces of unknown ID	Udzungwa Mts., Tanzania	X	
MCR14	R. Aba Nzenme	Faeces of unknown ID	Monts de Cristal, Gabon	X	
MCR34	J. L. Dew	Faeces of unknown ID	Monts de Cristal, Gabon	X	
MCR35	J.L. Dew	Faeces of unknown ID	Monts de Cristal, Gabon	X	
SA08	B. Curran	Faeces of unknown ID	Salonga, DRC	X	
SA15	B. Curran	Faeces of unknown ID	Salonga, DRC	X	
ND03	E. Stokes, F. Maisels	Faeces of unknown ID	Ndoki, Republic of the Congo	X	X
ND04	E. Stokes, F. Maisels	Faeces of unknown ID	Ndoki, Republic of the Congo		X
ND07	E. Stokes, F. Maisels	Faeces of unknown ID	Ndoki, Republic of the Congo		X
ND10	E. Stokes, F. Maisels	Faeces of unknown ID	Ndoki, Republic of the Congo	X	
ND12	E. Stokes, F. Maisels	Faeces of unknown ID	Ndoki, Republic of the Congo	X	
Lope 25	P. Mickala	Faeces of unknown ID	Lope, Gabon		X
Lope 36	P. Mickala	Faeces of unknown ID	Lope, Gabon		X
Lope30	P. Mickala	Faeces of unknown ID	Lope, Gabon		X
Lope101	P. Mickala	Faeces of unknown ID	Lope, Gabon	X	
Lope102	P. Mickala	Faeces of unknown ID	Lope, Gabon	X	
Oss02	P. Mickala	Faeces of unknown ID	Ossele, Gabon		X
Oss06	P. Mickala	Faeces of unknown ID	Ossele, Gabon		X
Oss07	P. Mickala	Faeces of unknown ID	Ossele, Gabon		X
Oss10	P. Mickala	Faeces of unknown ID	Osselle, Gabon		X
Kess10	P. Mickala	Faeces of unknown ID	Kessala, Gabon		X
IP94	P. Mickala	Faeces of unknown ID	Ipassa, Gabon		X
IP100	P. Mickala	Faeces of unknown ID	Ipassa, Gabon		X

X = not submitted

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.

Species	*Sfc*I	*Eco*RV	*Taq*I	*Bam*HI	*Bfu*AI	*Sty*I	*Bst*API
C. monticola	(188, 365) 7/7* (338, 215) 2/7	0/7	(242, 311) 5/7	(73, 480) 4/7	(356, 197) 1/7 (512, 41) 1/7	(447, 106) 5/7	(335, 218) 6/7
C. nigrifrons	0/3	0/3	0/3	(370, 183) 3/3	(356, 197) 3/3	(447, 106) 3/3	(335, 218) 3/3
C. ogilbyi	0/1	(444, 109) 1/1	(242, 311) 1/1	(73, 480) 1/1	(512, 41) 1/1	(447, 106) 1/1	0/1
C. callipygus	0/6	(444, 109) 6/6	(242, 311) 6/6	(73, 480) 6/6	(356, 197) 6/6 (512, 41) 6/6	(447, 106) 6/6	(335, 218) 6/6
C. leucogaster	0/2	0/2	(242, 311) 2/2	(73, 480) 1/2 (370, 183) 2/2	(356, 197) 2/2	(447, 106) 2/2 (102, 451) 1/2 (77, 476) 1/2	(335, 218) 2/2
C. silvicultor	0/4	0/4	0/4	(73, 480) 4/4 (370, 183) 4/4	0/4	(447, 106) 4/4 (338, 215) 1/4	0/4
C. dorsalis	0/8	0/8	(242, 311) 1/8	(370, 183) 8/8	0/8	(447, 106) 8/8	0/8
T. spekei	0/2	0/2	0/2	(73, 480) 2/2 (370, 183) 2/2	0/2	(338, 215) 2/2	0/2
T. scriptus	(134, 419) 4/4	0/2	(288, 265) 4/4 (242, 311) 4/4	(370, 183) 4/4	0/4	(507, 46) 1/4	0/4
H. aquaticus	(188, 365) 2/2	0/2	(242, 311) 1/2	(73, 480) 2/2 (370, 183) 2/2	0/2	(447, 106) 2/2 (507, 46) 2/2	0/2

* Fractions in the table illustrate the number of individual samples (numerator) that have the RFLP cut site for a given restriction enzyme assessed over the total number of samples for a given species (denominator). Restriction profiles in grey are uninformative.

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

Species	*Ssp*I	*BsrG*I	*Nci*I	*Bsr*I	*Sty*I	*Nde*I	*Fsp*I	*Psi*I
C. *monticola* (clade A)	**(250, 400) 4/5***	0/5	0/5	0/5	(135, 515) 1/5	0/5	0/5	(295, 345) 1/5
C. *callipygus* (clade A)	0/7	**(70, 240, 330) 6/7 (70, 570) 1/7**	0/7	0/7	0/7	(215, 425) 1/7 (245, 395) 1/7	0/7	(110, 250, 280) 5/7 (250, 390) 2/7
C. *callipygus* (clade B)	0/2	0/2	**(265, 350) 2/2**	0/2	0/2	0/2	0/2	0/2
T. spekei	0/6	0/6	0/6	**(215, 435) 6/6**	(245, 405) 3/6 (60, 185, 405) 3/6	0/6	0/6	0/6
T. *scriptus*	0/5	0/5	0/5	0/5	**(185, 470) 2/5 (185, 220, 250) 2/5 (135, 520) 1/5**	0/5	0/5	(80, 575) 1/5
C. *dorsalis*	0/21	0/21	0/21	0/21	0/21	**(225, 420) 15/21**	0/21	0/21
C. *leucogaster*	0/5	0/5	0/5	0/5	0/5	0/5	**(210, 430) 4/5**	(200, 440) 3/5
C. *nigrifons* (clade A)	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3
C. *nigrifons* (clade B)	0/2	0/2	0/2	0/2	0/2	0/2	0/2	**(80, 190, 330) 2/2**
C. *silvicultor*	0/7	0/7	0/7	0/7	0/7	(205, 435) 1/7	0/7	0/7

* Fractions in the table illustrate the number of individual samples (numerator) that have the RFLP cut site for a given restriction enzyme assessed over the total number of samples for a given species (denominator). Restriction profiles in BOLD are informative.

Chapter 4 appendices

Table S1. Sample codes, sample collectors, species identities, and geographic origins of samples used in the present study.

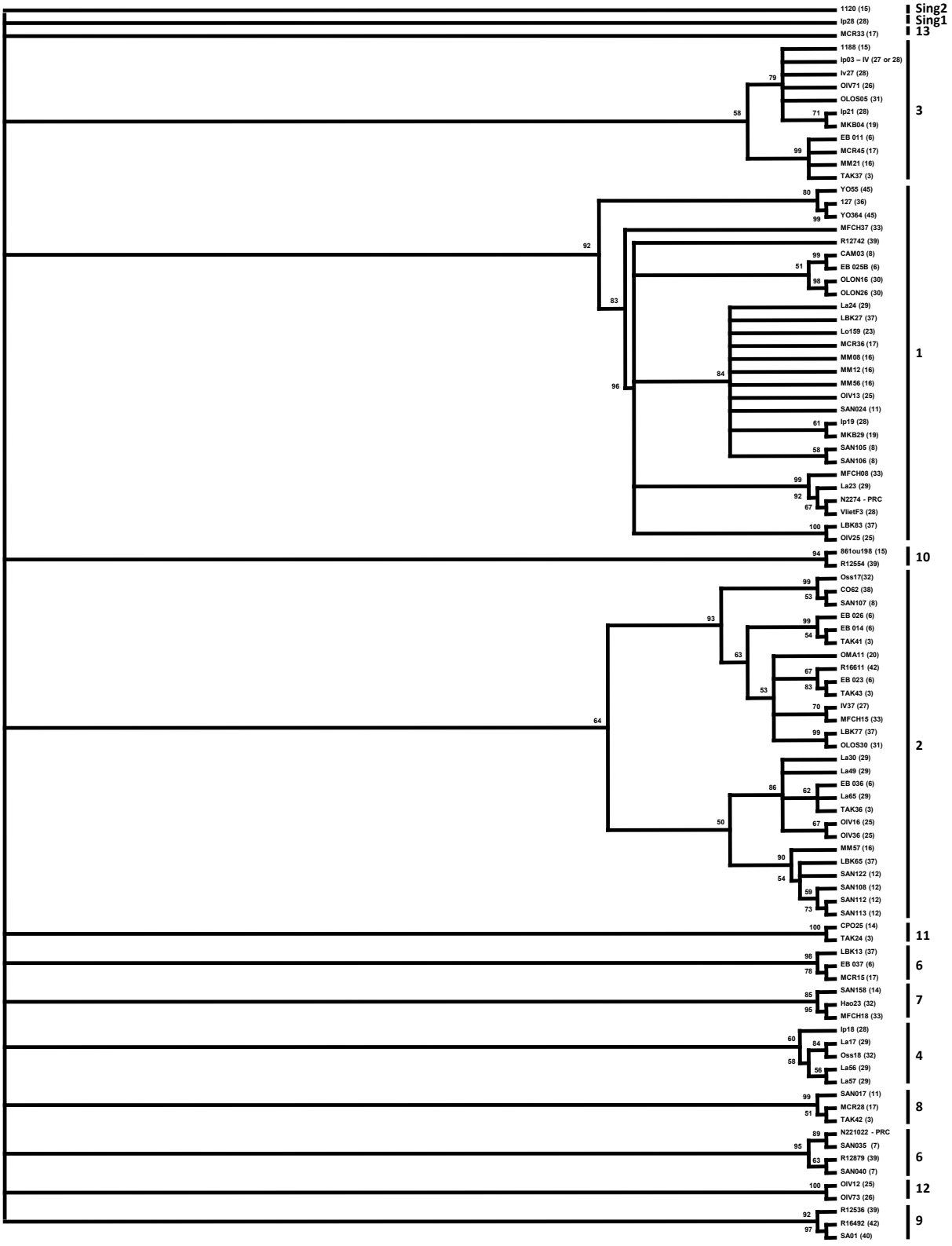
Sample code	Provider	Species	Geographic origin
DP4	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP5	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP6	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP19	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP20	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP43	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP44	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP45	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP46	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP47	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP48	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP49	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP60	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP62	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP67	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP66	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP54	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP55	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP56	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP57	Deborah Pires, Mitchell J. Eaton	*C. monticola*	Ndoki, Republic of Congo
DP10	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP11	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP12	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP21	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP22	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP23	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP24	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP25	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP26	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP27	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP72	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP73	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP74	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP75	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP76	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP77	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP78	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP79	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo

Table S1 (continued). Sample codes, sample collectors, species identities, and geographic origins of samples used in the present study.

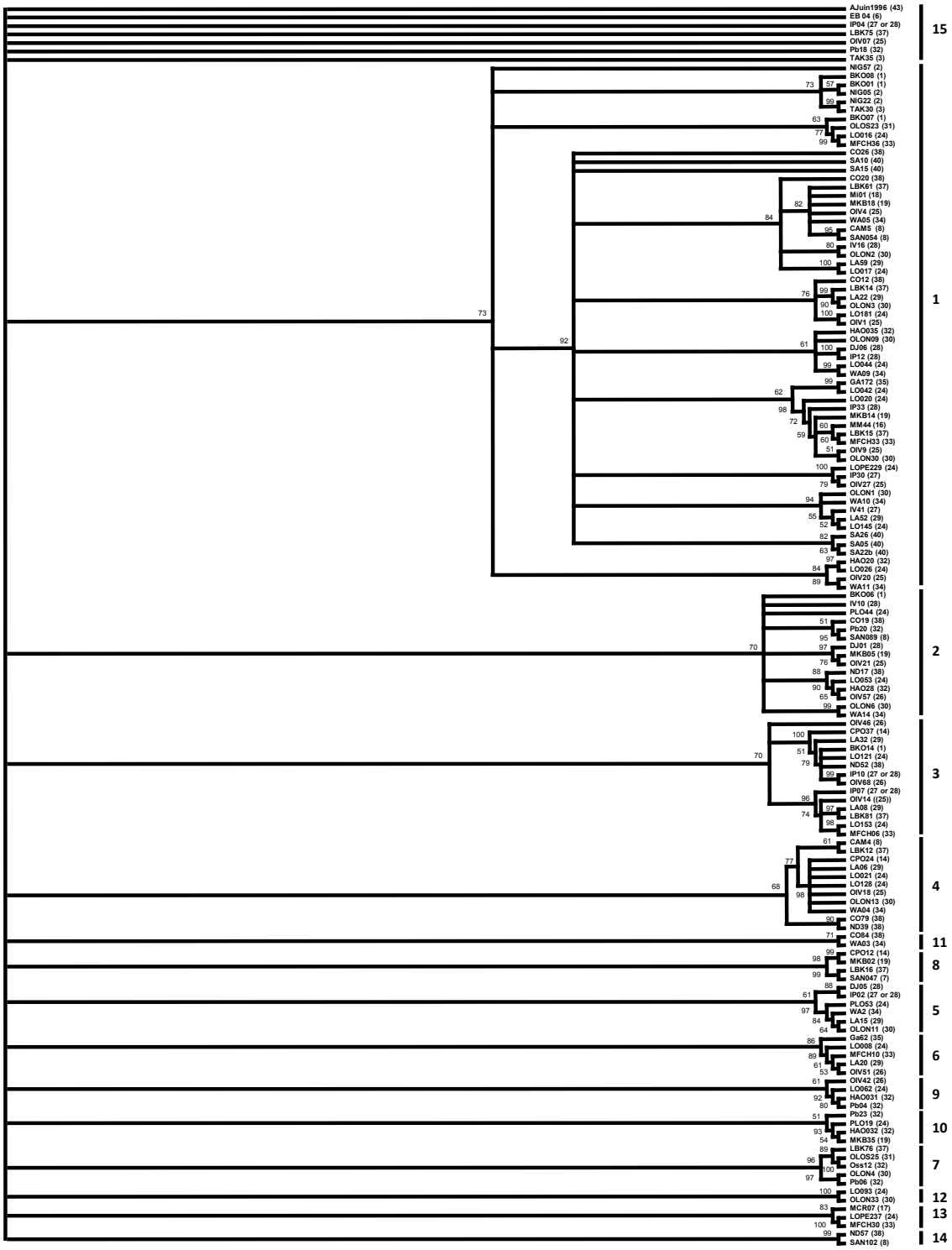
DP80	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP81	Deborah Pires, Mitchell J. Eaton	*C. callipygus*	Ndoki, Republic of Congo
DP01	Deborah Pires, Mitchell J. Eaton	*C. silvicultor*	Ndoki, Republic of Congo
DP02	Deborah Pires, Mitchell J. Eaton	*C. silvicultor*	Ndoki, Republic of Congo
DKME27	Stevens Touladjan	*C. silvicultor*	Okondja, Gabon
DP42	Deborah Pires, Mitchell J. Eaton	*C. silvicultor*	Ndoki, Republic of Congo
DKME23	Stevens Touladjan	*C. silvicultor*	Okondja, Gabon
DKME42	Stevens Touladjan	*C. silvicultor*	Okondja, Gabon
DP07	Deborah Pires, Mitchell J. Eaton	*C. nigrifrons*	Ndoki, Republic of Congo
DP08	Deborah Pires, Mitchell J. Eaton	*C. nigrifrons*	Ndoki, Republic of Congo
DP09	Deborah Pires, Mitchell J. Eaton	*C. nigrifrons*	Ndoki, Republic of Congo
DP68	Deborah Pires, Mitchell J. Eaton	*C. nigrifrons*	Ndoki, Republic of Congo
DP69	Deborah Pires, Mitchell J. Eaton	*C. nigrifrons*	Ndoki, Republic of Congo
DP70	Deborah Pires, Mitchell J. Eaton	*C. nigrifrons*	Ndoki, Republic of Congo
DP13	Deborah Pires, Mitchell J. Eaton	*C. dorsalis*	Ndoki, Republic of Congo
DP15	Deborah Pires, Mitchell J. Eaton	*C. dorsalis*	Ndoki, Republic of Congo
DP35	Deborah Pires, Mitchell J. Eaton	*C. dorsalis*	Ndoki, Republic of Congo
DP36	Deborah Pires, Mitchell J. Eaton	*C. dorsalis*	Ndoki, Republic of Congo
DP37	Deborah Pires, Mitchell J. Eaton	*C. dorsalis*	Ndoki, Republic of Congo
DP87	Deborah Pires, Mitchell J. Eaton	*C. dorsalis*	Ndoki, Republic of Congo
DP17	Deborah Pires, Mitchell J. Eaton	*C. leucogaster*	Ndoki, Republic of Congo
DP28	Deborah Pires, Mitchell J. Eaton	*C. leucogaster*	Ndoki, Republic of Congo
DP29	Deborah Pires, Mitchell J. Eaton	*C. leucogaster*	Ndoki, Republic of Congo
DP30	Deborah Pires, Mitchell J. Eaton	*C. leucogaster*	Ndoki, Republic of Congo
DP32	Deborah Pires, Mitchell J. Eaton	*C. leucogaster*	Ndoki, Republic of Congo
DP34	Deborah Pires, Mitchell J. Eaton	*C. leucogaster*	Ndoki, Republic of Congo
DKME52	Stevens Touladjan	*T. spekei*	Okondja, Gabon
FR07	Stevens Touladjan	*T. spekei*	Franceville, Gabon
FR120	Stevens Touladjan	*T. spekei*	Franceville, Gabon
Lam01	Stevens Touladjan	*T. spekei*	Lambaréné, Gabon
Lam06	Stevens Touladjan	*T. spekei*	Lambaréné, Gabon
Lam07	Stevens Touladjan	*T. spekei*	Lambaréné, Gabon
Lam03	Stevens Touladjan	*H. aquaticus*	Lambaréné, Gabon
OK15	Stevens Touladjan	*H. aquaticus*	Okondja, Gabon
FR05	Stevens Touladjan	*H. aquaticus*	Franceville, Gabon
FR16	Stevens Touladjan	*H. aquaticus*	Franceville, Gabon
FR19	Stevens Touladjan	*H. aquaticus*	Franceville, Gabon
OK19	Stevens Touladjan	*H. aquaticus*	Okondja, Gabon

Chapter 5 appendices

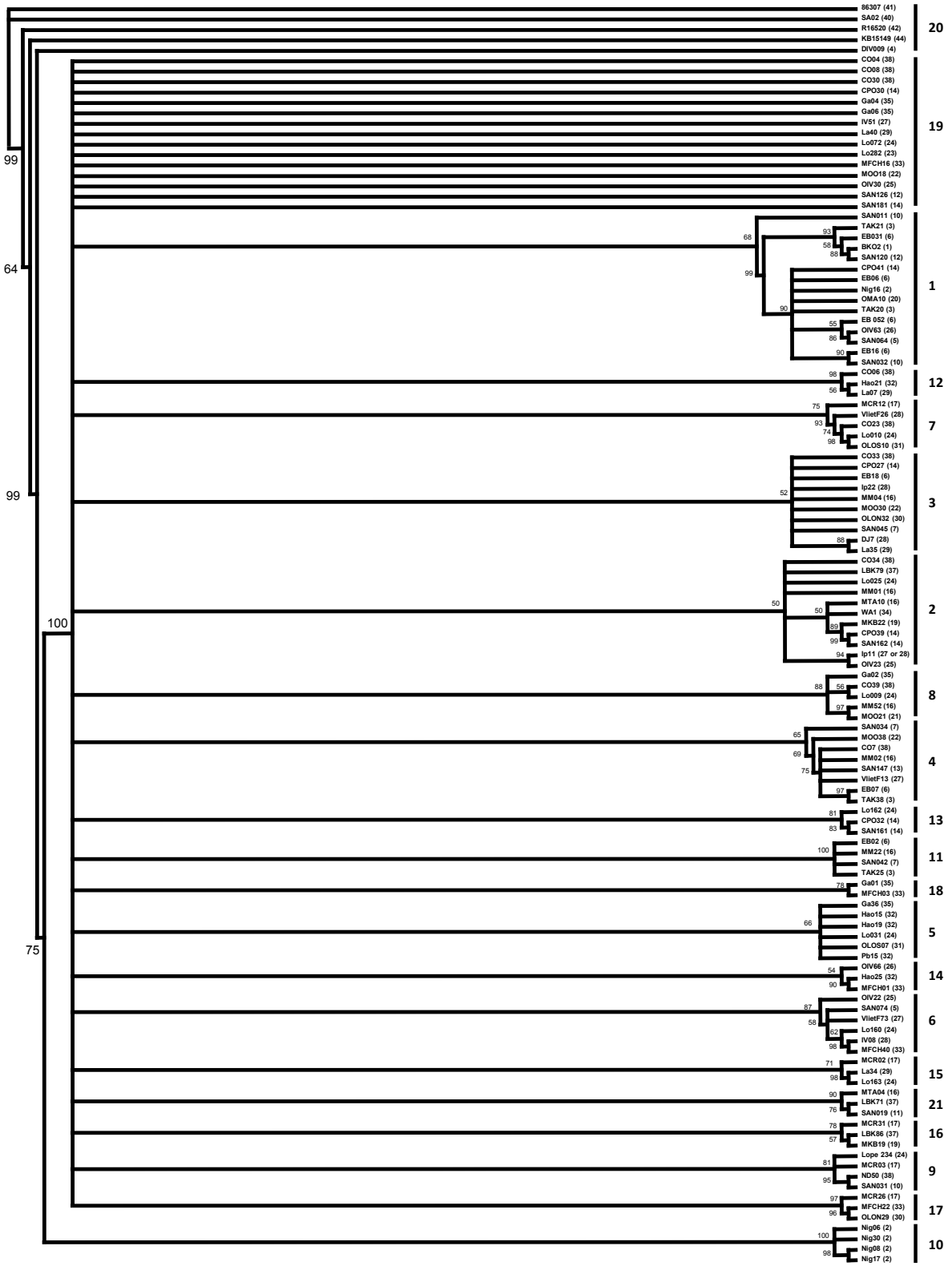
Supplementary Figure 1. *Cephalophus dorsalis* control region neighbor joining phylogenetic tree with bootstrap values based on 1000 replications using the nucleotide substitution model HKY+I+ G. Numbers on the right side refer to the haplogroups illustrated in the map of the study area (Figure 2a). Numbers in parentheses refer to collection sites illustrated in the map of the study area (Figure 1).



Supplementary Figure 2. *Cephalophus callipygus* control region neighbor joining phylogenetic tree with bootstrap values based on 1000 replications using the nucleotide substitution model GTR+I+G. Numbers on the right side refer to the haplogroups illustrated in the map of the study area (Figure 2b). Numbers in parentheses refer to collection sites illustrated in the map of the study area (Figure 1).

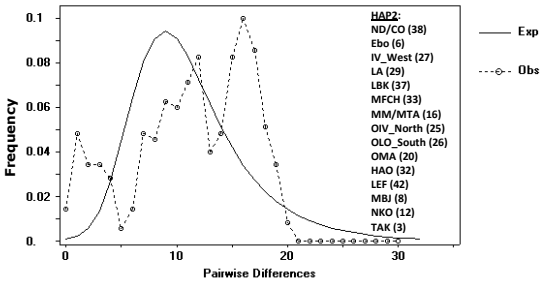


Supplementary Figure 3. *Philantomba monticola* control region neighbor joining phylogenetic tree with bootstrap values based on 1000 replications using the nucleotide substitution model TrN+ G (TAMNEI in Paup). Numbers on the right side refer to the haplogroups illustrated in the map of the study area (Figure 2c). Numbers in parentheses refer to collection sites illustrated in the map of the study area (Figure 1).

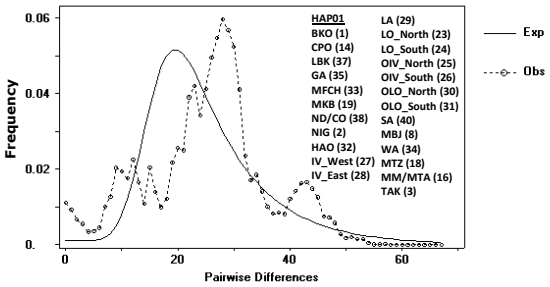


Supplementary Figure 4. Significant mismatch distributions of control region pairwise genetic differences in (a) *C. dorsalis*, (b) *C. callipygus* and (c) *P. monticola*. The solid and dashed lines are the respective expected and observed distributions of pairwise nucleotide comparisons based on a model of exponential population growth. The combined sites tested are listed from the top right of the figure. Numbers in parentheses refer to collection sites illustrated in the map of the study area (Figure 1).

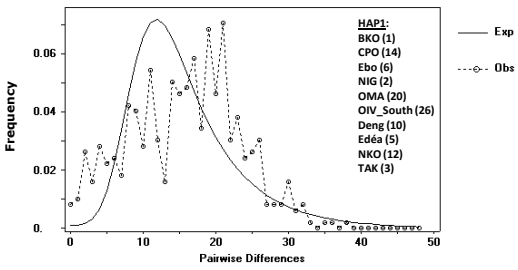
a)

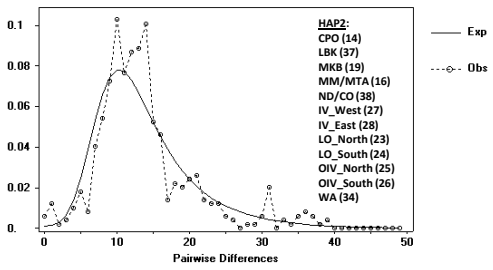


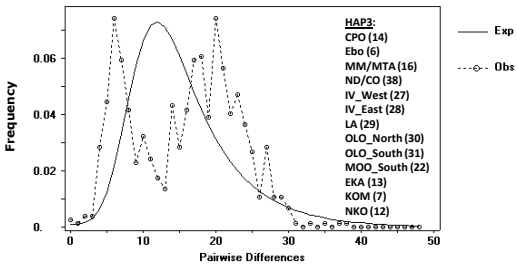
b)

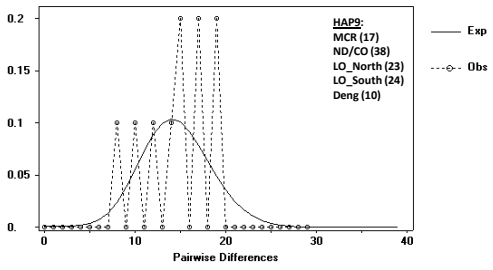


c)



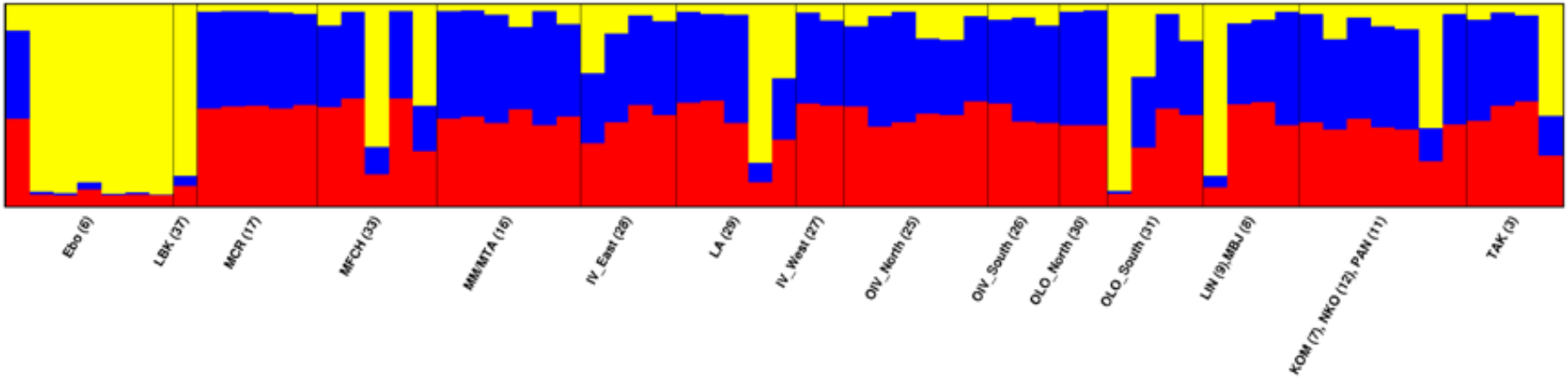




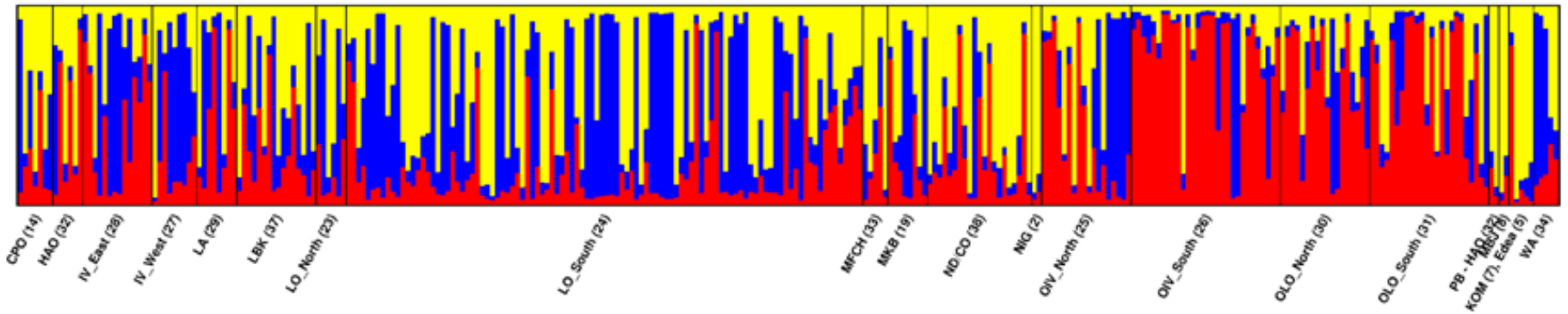


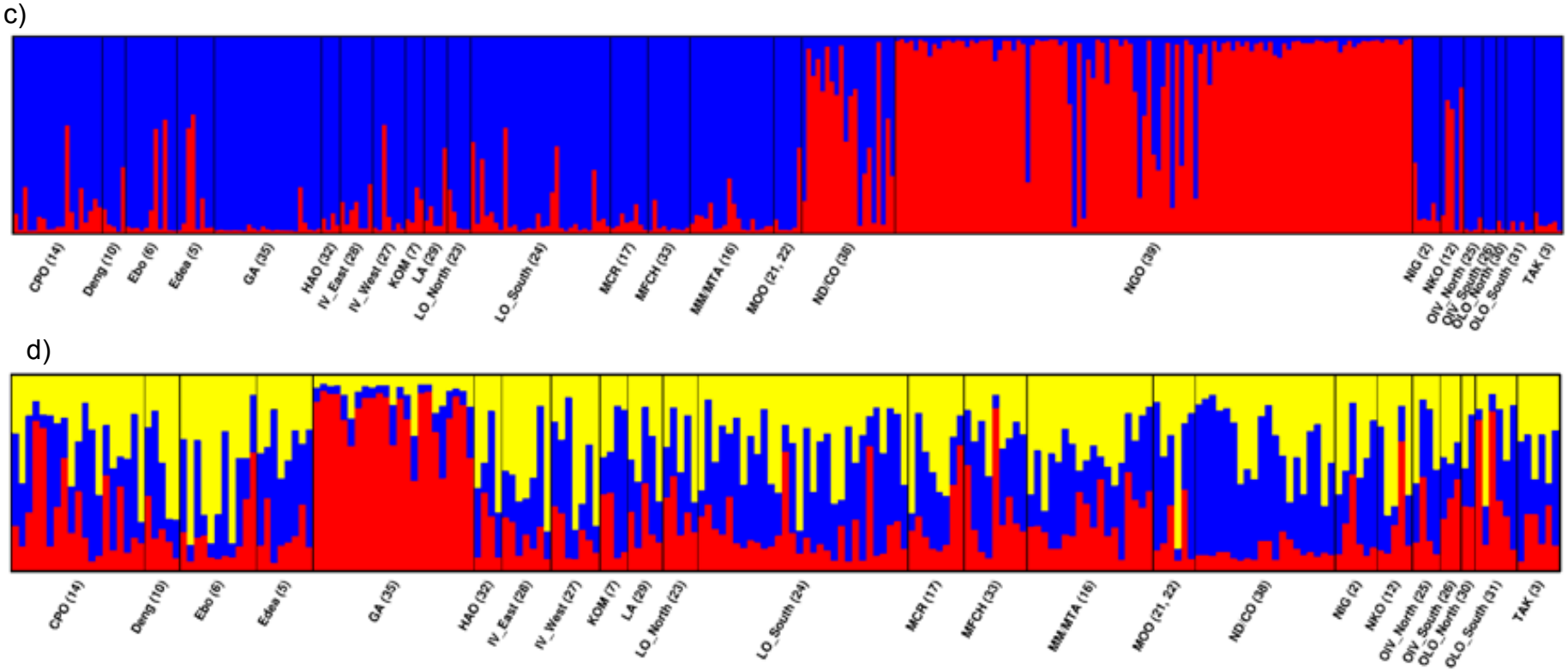
Supplementary Figure 5. Graphical representation of the STRUCTURE output of (a) *C. dorsalis* (b) *C. callipygus* and (c) *P. monticola* with first order relatives included and a best K of 3, 3, 2, respectively. For *P. monticola* (d) results are also shown with Ngotto excluded. These graphics represents proportional cluster membership of each individual genotype among populations and K groups. Numbers in parentheses refer to collection sites illustrated in the map of the study area (Figure 1).

a)



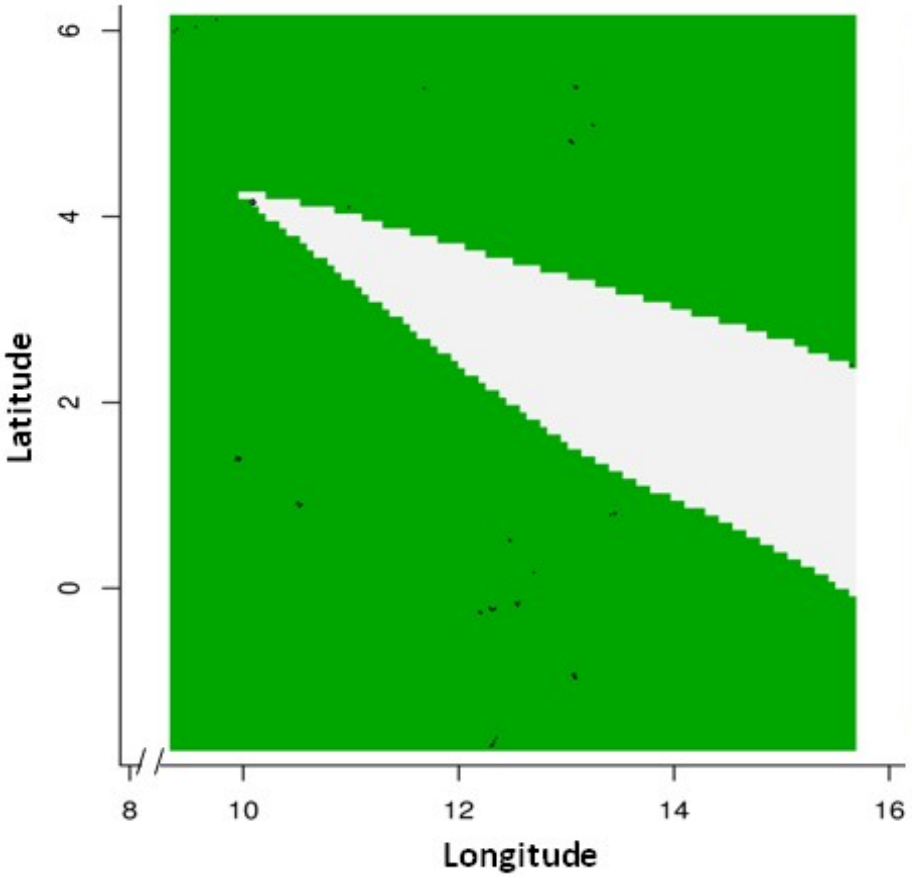
b)

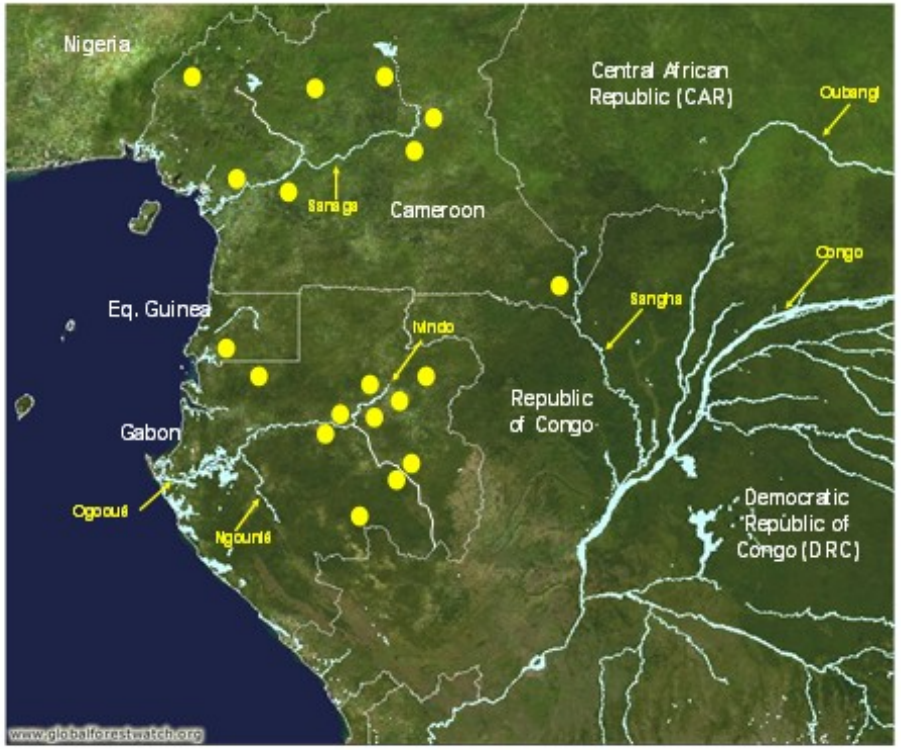




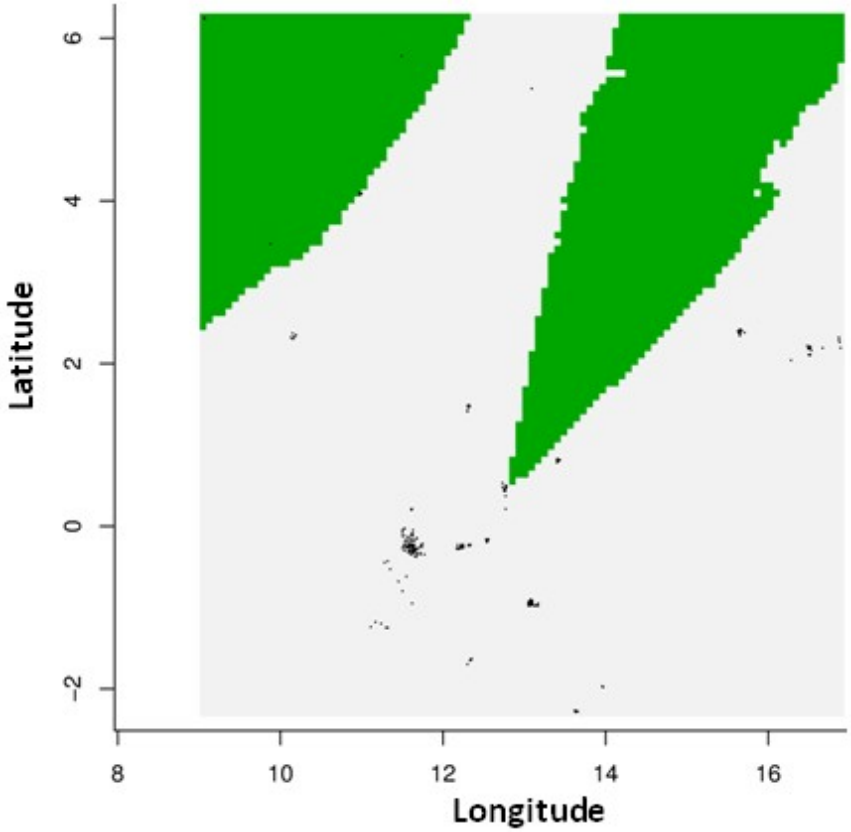
Supplementary Figure 6. Maps showing sampling locations in Central Africa of (a) *C. dorsalis* (b) *C. callipygus* and (c) *P. monticola* with first order relatives included. For *P. monticola* results are shown with all the sites except Ngotto (d). For each species, the map on the left is a synthetic map produced by GENELAND which shows the ranges of inferred genetic groups (green, white, and yellow). For each species, the map on the right indicates the approximate sampling locations (yellow dots) whereas the map on the left illustrates sampling locations with black dots. Geographic distances are not to scale in the map on the right.

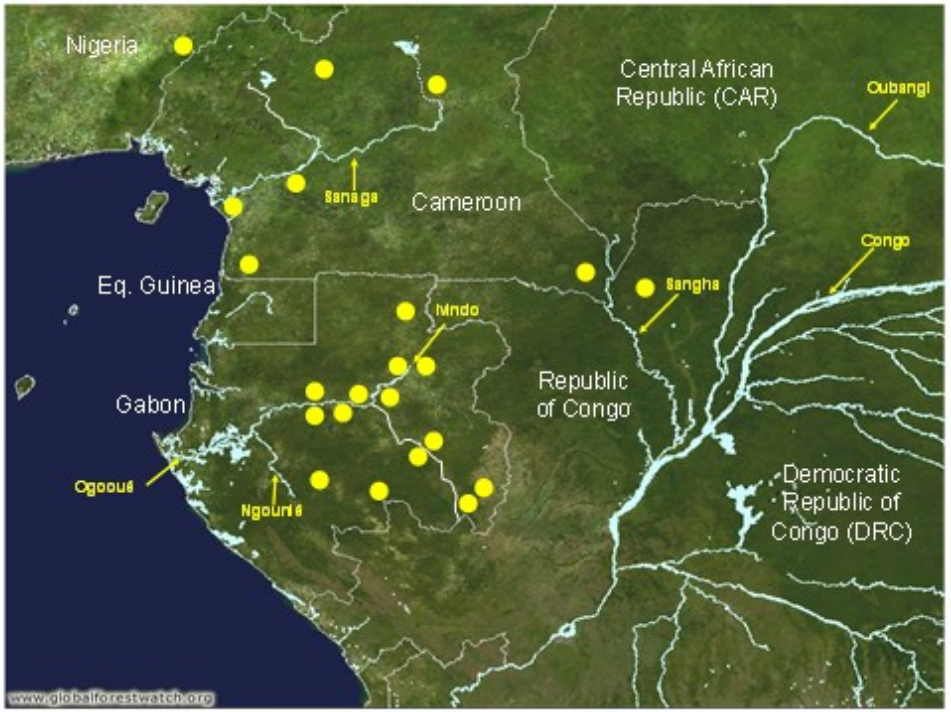
a)


Latitude
6
4
2
0
8
10
12
14
16
Longitude

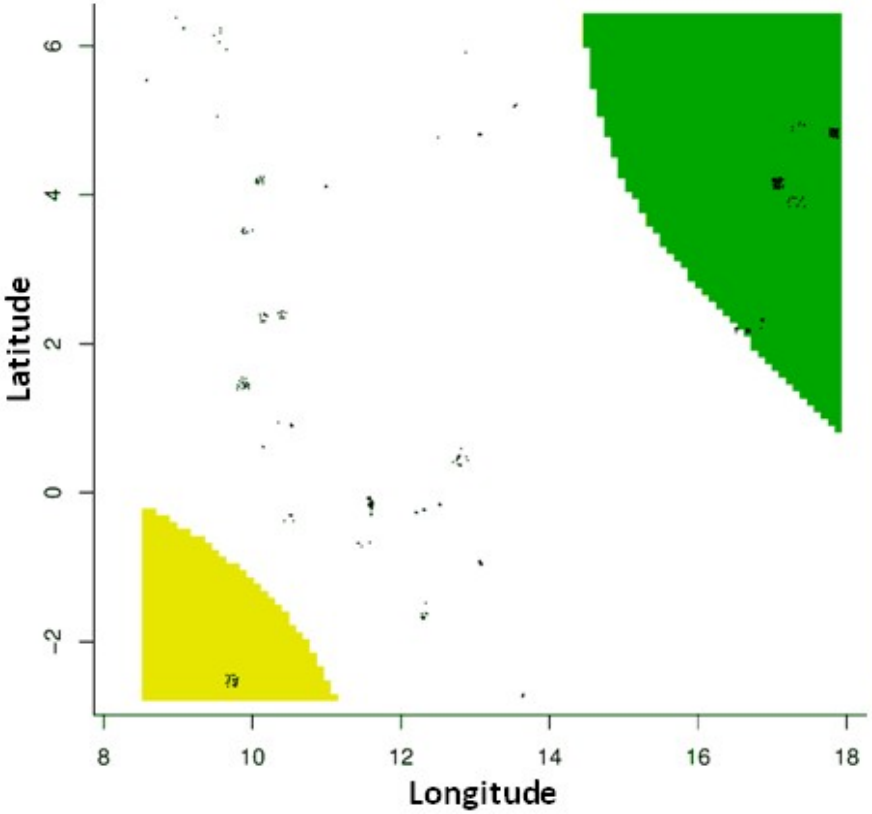

Nigeria
Central African Republic (CAR)
Oubangi
Sanaga
Cameroon
Congo
Sangha
Eq. Guinea
Mindo
Republic of Congo
Gabon
Ogooué
Ngounié
Democratic Republic of Congo (DRC)
www.globalforestwatch.org

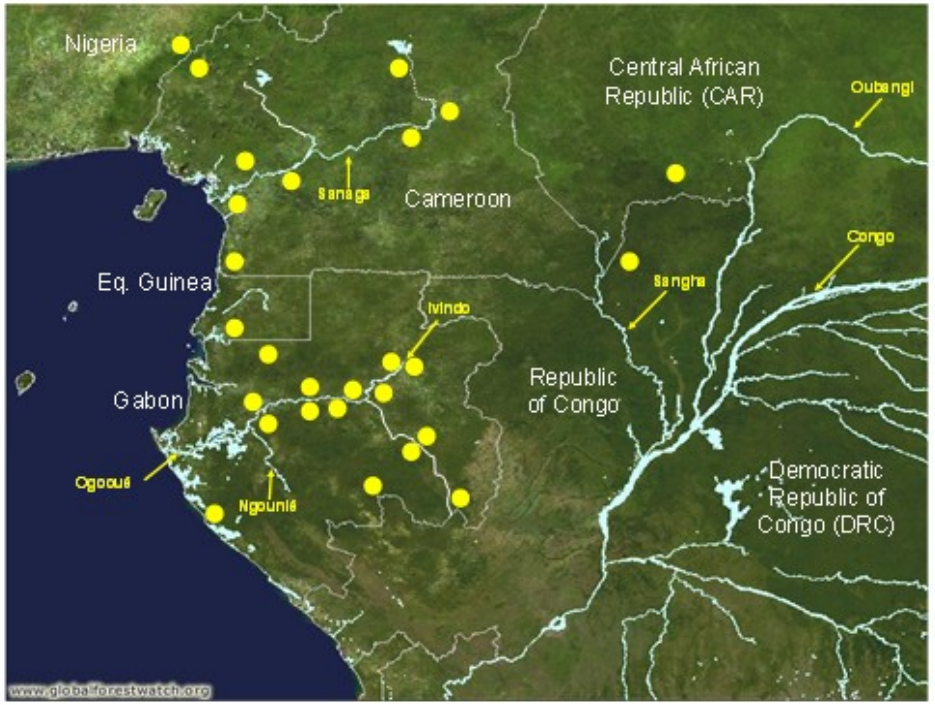
b)


Latitude
6
4
2
0
−2
8
10
12
14
16
Longitude

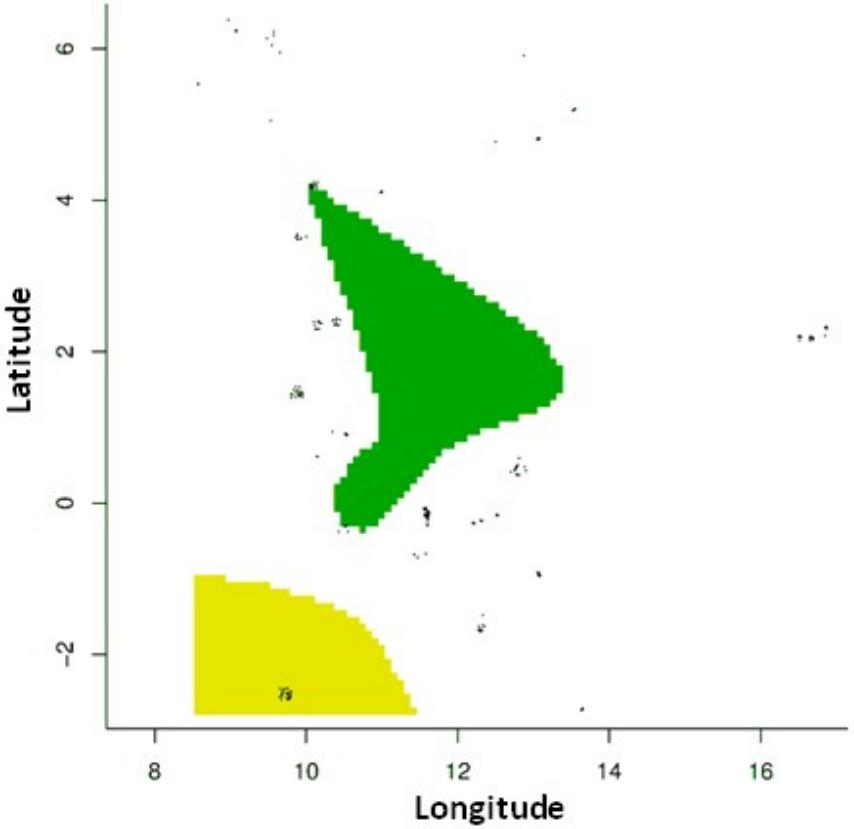

Nigeria
Central African Republic (CAR)
Oubangi
Sanaga
Cameroon
Congo
Eq. Guinea
Sangha
Mindo
Republic of Congo
Gabon
Ogooué
Ngounié
Democratic Republic of Congo (DRC)
www.globalforestwatch.org

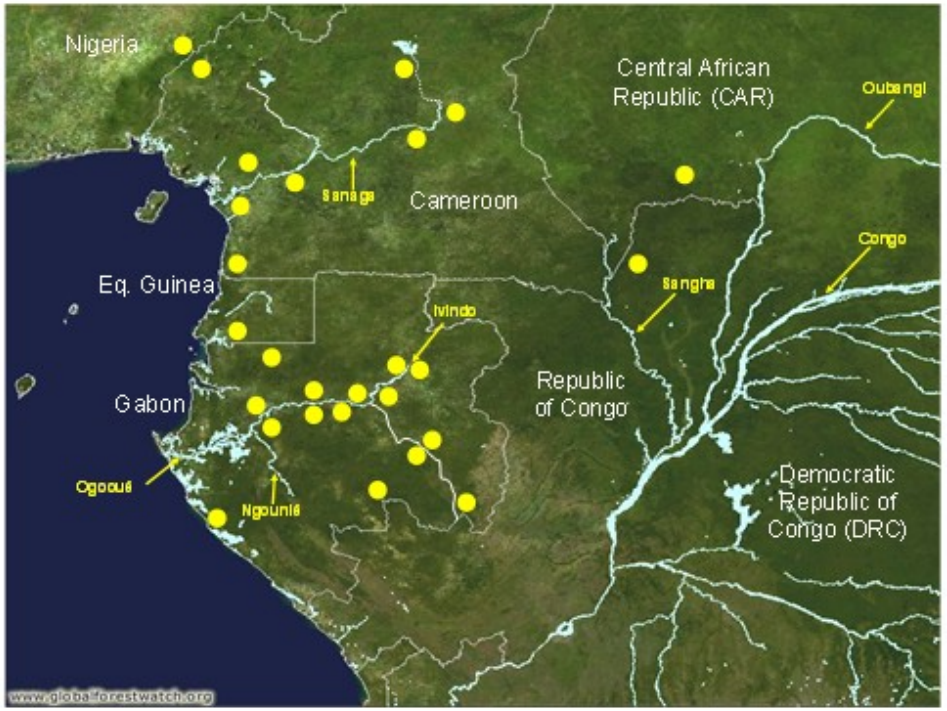
c)


Latitude
Longitude
6
4
2
0
−2
8
10
12
14
16
18


Nigeria
Central African Republic (CAR)
Oubangi
Sanaga
Cameroon
Congo
Eq. Guinea
Sangha
Ivindo
Republic of Congo
Gabon
Ogooué
Ngounié
Democratic Republic of Congo (DRC)
www.globalforestwatch.org

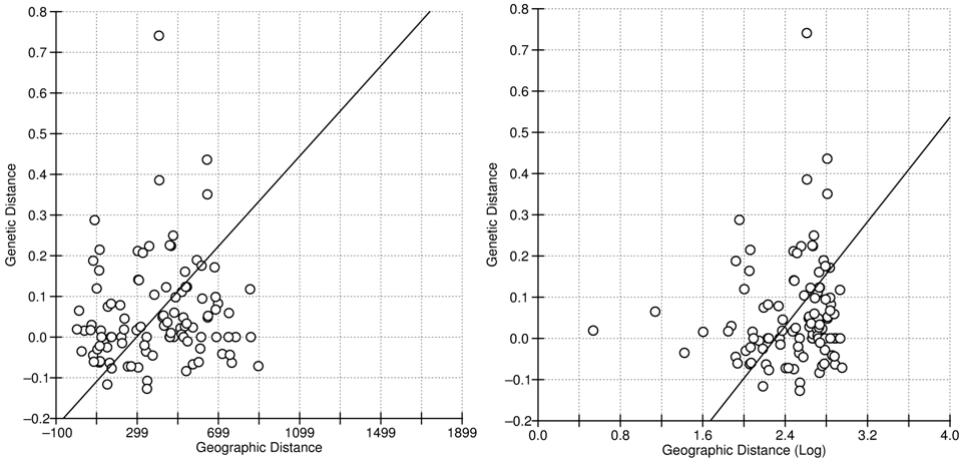
d)


Latitude
Longitude
6
4
2
0
-2
8
10
12
14
16

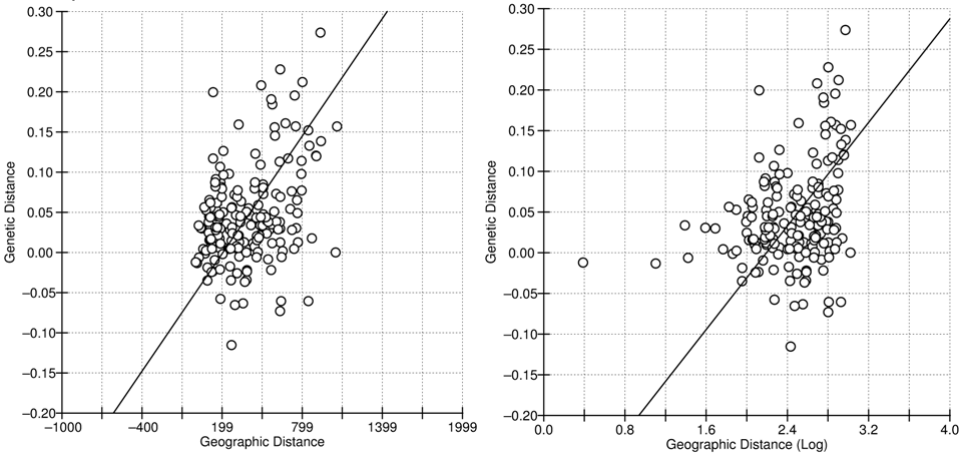

Nigeria
Central African Republic (CAR)
Oubangi
Sanaga
Cameroon
Congo
Sangha
Eq. Guinea
Ivindo
Gabon
Republic of Congo
Ogooué
Ngounié
Democratic Republic of Congo (DRC)
www.globalforestwatch.org

Supplementary Figure 7. Isolation by distance relationships plotted as Slatkin's linearized Fst [Fst/(1 – Fst)] against pairwise geographical distances in kilometers (left) and the logarithm of geographic distances (right) among (a) *C. dorsalis,* (b) *C. callipygus* and (c) *P. monticola* populations in Central Africa.

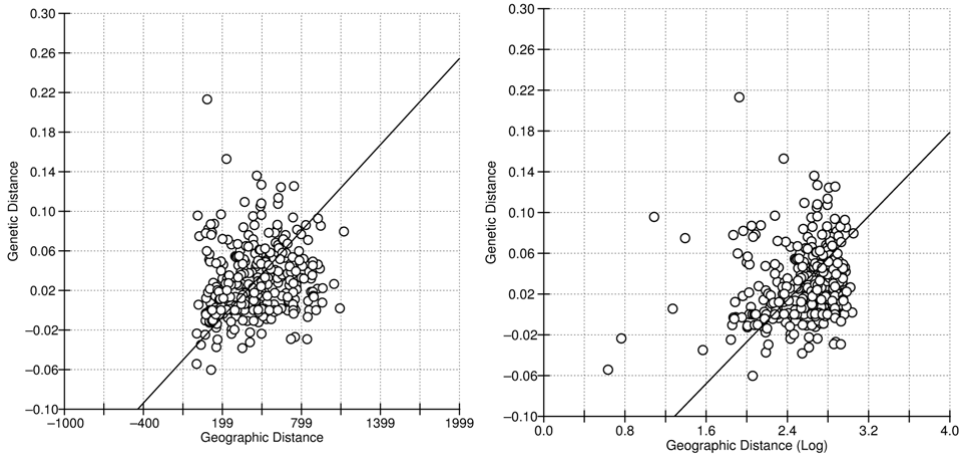
a)



b)



c)



Supplementary Table 1. Countries, site names, site numbers, letter codes, sample sizes, and sample origin of *C. dorsalis, C. callipygus*, and *P. monticola* as illustrated in Figure 1.

Country	Sites	Site numbers	Letter codes	Sample sizes	Sample origin
Eq. Guinea	Reserva Científica de la Caldera de Luba, Bioko	1	BKO	14	Feces
Nigeria	Cross River National Park	2	NIG	75	Feces
Cameroon	Takamanda Forest Reserve	3	TAK	44	Feces
Cameroon	Bamenda	4	DIV	1	Genebank
Cameroon	Douala-Edéa Forest Reserve/Ekoth	5	Edea	18	Feces
Cameroon	Ebo Forest	6	Ebo	54	Feces
Cameroon	Kombe	7	KOM	14	Feces
Cameroon	Parc National de Mbam et Djerem/Wouchaba	8	MBJ	68	Feces
Cameroon	Linté	9	LIN	7	Feces
Cameroon	Deng deng	10	DENG	9	Feces
Cameroon	Lom/Pangar	11	PAN	16	Feces
Cameroon	Nkolomaken	12	NKO	22	Feces
Cameroon	Ekanga/Masseng/Biwali/Ngamba Enduum	13	EKA	24	Feces
Cameroon	Parc National de Campo Ma'an	14	CPO	43	Feces
Cameroon	Dja Reserve	15	Dja	3	Genebank
Equatorial guinea	Monte Mitra/Monte Alen	16	MM/MTA	67	Feces
Gabon	Parc National des Monts de Cristal	17	MCR	56	Feces
Gabon	Midzic	18	MTZ	3	Feces
Gabon	Parc National de Minkébé	19	MKB	35	Feces
Gabon	Ogooué Maritime	20	OMA	15	Feces
Gabon	Moyen Ogooué (north and south)	21, 22	MOO	54	Feces
Gabon	Parc National de la Lopé (north and south)	23, 24	LO	354	Feces
Gabon	Ogooué Ivindo (north and south)	25, 26	OIV	80	Feces
Gabon	Ivindo/Ipassa/Dji dji (West and East)	27, 28	IV	101	Feces
Gabon	Langoué	29	LA	67	Feces
Gabon	Ogooué Lolo (north and south)	30, 31	OLO	79	Feces
Gabon	Boumango, Parc National des Plateaux Batéké, Léconi, Bakoumba, Ossélé/Kessala	32	HAO	102	Feces
Gabon	Massif du Chaillu	33	MFCH	40	Feces
Gabon	Parc National de Waka	34	WA	21	Feces
Gabon	Parc National de Loango-Sette Cama (including Malounga)	35	GA	67	Feces, Genebank

Supplementary Table 1 (continued). Countries, site names, site numbers, letter codes, sample sizes, and sample origin of *C. dorsalis, C. callipygus*, and *P. monticola* as illustrated in Figure 1.

Republic of Congo	Parc National d'Odzala	36	ODZ	1	Genebank
Cameroon	Parc National de Lobéké	37	LBK	110	Feces
Republic of Congo	Parc National de Nouabelé Ndoki	38	ND/CO	178	Feces
Central African Republic	Parc National de Ngotto	39	NGO	163	Feces, tissues, Genebank
Democratic Republic of Congo	Parc National de Salonga	40	SA	27	Feces
Democratic Republic of Congo	Kisangani	41	KIS	1	Genebank
Republic of Congo	Reserve de Lefini	42	LEF	3	Genebank
Republic of Congo	Brazzaville	43	BRA	1	Genebank
South Africa	Cape province	44	CAP	1	Genebank
Guinae	Diecke	45	DIE	2	Genebank
Gabon	Parc National de Loango-Iguela	None	IGUE	34	Feces
Total				**2074**	

Supplementary Table 2. *C. dorsalis* SAMOVA summary statistics of FCT and P-value (K=2 to K=10).

K	2	3	4	5	6	7	8	9	10
% variation	18.50	17.68	13.31	13.09	12.81	13.40	14.19	14.79	15.19
P-value	0.0694	0.0166	<10-5	<10-5	<10-5	<10-5	<10-5	<10-5	<10-5
Groups	12, Rest	12, (8, 9), Rest	12, (8, 9), (6, 3), Rest	12, (8, 9), (6, 3), 29, Rest	12, (8, 9, 16), (6, 3), (25, 26, 24), 29, Rest	(8, 9), (6, 3), (37, 36, 38, 33, 30, 31), (27, 28, 17, 16), 29, 12, (39, 25, 26, 24)	(8, 9), (27, 28, 17, 16), 12, (37, 36, 38, 33, 30, 31), (25, 26, 24) 39, 29, (6, 3)	(8, 9), (27, 28, 17), 16, 12, (37, 36, 38, 33, 30, 31), (25, 26, 24), 39, 29, (6, 3)	(8, 9), (30, 31), (27, 28, 17), (37, 36, 38, 33), 16, (25, 26, 24), 39, (29), (6, 3)

Supplementary Table 3. *C. callipygus* SAMOVA summary statistics of FCT and P-value (K=2 to K=10).

K	2	3	4	5	6	7	8	9	10
% variation	19.52	18.25	16.07	13.88	12.02	11.33	10.74	10.16	9.63
FCT P-value	0.04008	0.00782	<10-5	<10-5	0.00098	<10-5	<10-5	<10-5	<10-5
Groups	(2, 3), Rest	(2, 3), 40, Rest	(2, 3), 40, 1, Rest	(2, 3), 40, 1, 23, Rest	(2, 3), 40, 1, 23, 38, Rest	(2, 3), 40, 1, 23, 38, 19, Rest	(2, 3), 40, 1, 23, 38, 19, 14, Rest	(2, 3), 40, 1, 23, 38, 19, 14, 8, Rest	(2, 3), 40, 1, 23, 38, 19, 14, 8, 32, Rest

Supplementary Table 4. *P. monticola* SAMOVA summary statistics of FCT and P-value (K=2 to K=10, without Nigerian samples).

K	2	3	4	5	6	7	8	9	10
% variation	40.57	35.24	34.1	33.86	33.24	32.41	29.62	28.17	27.24
FCT P-value	0.02933	0.00293	<10-5	<10-5	<10-5	<10-5	<10-5	<10-5	<10-5
	5, rest	5, 32, Rest	5, 32, (3, 4), Rest	5, (6, 3, 4), 7, 32, Rest	5, (6, 3, 4), 7, 32, 10, Rest	5, 6, (3, 4), 7, 10, 32, Rest	5, (6, 3, 4), 7, 32, 10, 33, 19, Rest	10, 35, 6, 7, (21, 22), 5, (3, 4), 32, Rest	19, 6, 10, (21, 22), (3, 4), 5, 35, 32, 7, Rest

Supplementary Table 5. The location of *C. dorsalis, C. callipygus*, and *P. monticola* haplogroups by site as illustrated in Figure 2. Number of individuals for each haplogroup is indicated in parenthesis.

Sites names	***C. dorsalis***	***C. callipygus***	***P. monticola***
DIV (4)			Sing2 (1)
BKO (1)		1 (3), 2 (1), 3 (1)	1 (2)
BRA (43)		15 (1)	
CPO (14)	7 (1), 11 (1)	1 (6), 3 (1), 4 (2), 5 (1), 8 (2)	1 (1), 2 (4), 3 (10), 4 (1), 13 (3), Sing1 (3)
CAP (44)			Sing2 (1)
Deng (10)			1 (2), 9 (1)
DIE (45)	1 (2)		
Dja (15)	3 (1), 10 (1), 15 (1)		
Ebo (6)	1 (1), 2 (4), 3 (2), 6 (2)	15 (1)	1 (10), 3 (1), 4 (3), 11 (5)
Edea (5)			1 (7), 8 (1)
EKA (13)			3 (1), 4 (2)
GA (35)		1 (1), 6 (1)	2 (3), 5 (1), 8 (3), 18 (3), Sing1 (26)
HAO (32)	2 (1), 4 (1), 7 (1)	1 (2), 2 (3), 7 (2), 9 (3), 10 (3), 15 (2)	5 (5), 12 (1), 14 (1)
IV_East (28)	1 (2), 3 (2), 4 (1)	1 (5), 2 (4), 3 (4), 5 (1), 15 (3)	2 (1), 3 (5), 6 (2), 7 (1), Sing1 (3)
IV_West (27)	1 (1), 2 (1), 14 (1)	1 (6), 2 (3), 3 (3), 4 (1), 5 (1), 8 (1)	2 (3), 3 (1), 4 (1), 6 (1), Sing1 (1)
KIS (41)			Sing2 (1)
KOM (7)	5 (2)	8 (1)	3 (1), 4 (2), 11 (2)
LA (29)	1 (2), 2 (4), 4 (3)	1 (7), 3 (9), 4 (3), 5 (2), 6 (2)	3 (3), 8 (1), 12 (1), 15 (1), Sing1 (4)
LEF (42)	2 (1), 9 (1)		Sing2 (1)
LIN (9)	1 (1)		
LBK (37)	1 (2), 2 (2), 6 (1)	1 (5), 3 (2), 4 (4), 7 (2), 8 (5), 15 (1)	2 (1), 16 (2), 21 (1), Sing1 (1)
LO_North (23)		1 (2), 4 (5), 13 (1)	2 (3), 8 (2), 9 (1), Sing1 (2)

Supplementary Table 5 (continued). The location of *C. dorsalis, C. callipygus*, and *P. monticola* haplogroups by site as illustrated in Figure 2. Number of individuals for each haplogroup is indicated in parenthesis.

LO_South (24)	2 (1)	1 (60), 2 (20), 3 (9), 4 (14), 5 (1), 6 (15), 7 (1), 9 (11), 10 (2), 12 (1), 13 (1)	2 (7), 5 (7), 6 (1), 7 (2), 8 (7), 9 (2), 13 (1), 15 (4), 21 (5), Sing1 (5)
MFCH (33)	1 (2), 2 (2), 7 (1)	1 (3), 3 (1), 6 (3), 13 (1)	6 (1), 14 (1), 17 (1), 18 (5), 21 (2), Sing1 (2)
MBJ (8)	1 (3), 2(1)	1 (2), 2 (1), 4 (1), 14 (1)	
MKB (19)	1 (1), 3 (2)	1 (3), 2 (6), 8 (1), 10 (1)	2 (1), 16 (3), Sing1 (2)
MTZ (18)		1 (1)	
MCR (17)	1 (5), 3 (1), 6 (1), 8 (2), 13 (3)	13 (1)	4 (2), 7 (3), 9 (1), 15 (5), 16 (3), 17 (2), 21 (4)
MM/MTA (16)	1 (4), 2 (1), 3 (1)	1 (1)	2 (4), 3 (26), 4 (1), 6 (1), 8 (1), 11 (2), 21 (1), Sing1 (8)
MOO_North (21)			8 (1)
MOO_South (22)			3 (1), 4 (3), Sing1 (1)
ND/CO (38)	2 (1)	1 (45), 2 (4), 3 (1), 4 (3), 11 (5), 14 (3)	2 (10), 3 (1), 4 (2), 7 (4), 8 (1), 9 (1), 12 (3), Sing1 (8)
NGO (39)	1 (1), 2 (5), 9 (1), 10 (1)		
NIG (2)		1 (3), 15 (3)	1 (2), 10 (33)
NKO (12)	2 (5)		1 (1), 3 (1), 11 (1), 21 (1), Sing1 (2)
ODZ (36)	1 (1)		
OIV_North (25)	1 (3), 2 (2), 12 (1)	1 (12), 2 (3), 3 (2), 4 (1), 15 (2)	2 (2), 6 (1), 21 (2), Sing1 (3)
OIV_South (26)	3 (1), 12 (2)	1 (7), 2 (6), 3 (8), 4 (1), 6 (10), 9 (2)	1 (1), 2 (1), 14 (1), 17 (1), 21 (1)
OLO_North (30)	1 (2)	1 (8), 2 (5), 4 (3), 5 (3), 7 (3), 12 (1)	3 (1), 13 (1), 14 (1), 17 (1), Sing1 (1)
OLO_South (31)	2 (3), 3 (1)	1 (14), 2 (4), 5 (1), 7 (3), 12 (1)	3 (2), 5 (6), 7 (1), 14 (1), 21 (5), Sing1 (1)
OMA (20)	2 (1)		1 (1)
PAN (11)	1 (8), 8 (1)		21 (1)
SA (40)	9 (1)	1 (5)	Sing2 (1)
TAK (3)	2 (3), 3 (1), 8 (1), 11 (1)	1 (1), 15 (1)	1 (7), 4 (1), 11 (2)
WA (34)		1 (4), 2 (1), 4 (1), 5 (1), 8 (1), 11 (4)	2 (1)

Supplementary Table 6. *C. dorsalis* expected heterozygosity and Hardy-Weinberg Equilibrium P-values.

		He		Hardy-Weinberg Equilibrium test									
	n	**Mean**	**SD**	**BM2113**	**INRA40**	**BM1225**	**BM143**	**INRA05**	**BM1862**	**BM864**	**MM12**	**BM121**	**BM848**	**SR12**
Ebo (6)	6	0.7694	0.1211	0.06588	0.8598	0.2842	MONO	MONO	MONO	MONO	1	0.0597	0.7715	0.393
IV_East (28)	4	0.7242	0.2307	1	1	0.65786	MONO	MONO	MONO	MONO	0.1977	1	0.1972	1
LA (29)	5	0.7369	0.1605	1	0.2326	0.69085	MONO	MONO	MONO	MONO	0.4296	0.4664	0.6006	0.4284
MM/MTA (16)	5	0.6474	0.2593	1	1	0.65679	MONO	MONO	MONO	MONO	1	0.0284	0.6222	1
OIV_North (25)	5	0.7169	0.1707	0.33446	1	0.31677	0.85146	1	1	0.59982	1	0.1913	0.849	1
OLO_South (31)	4	0.7693	0.2399	1	1	MONO	1	0.33415	1	1	0.4706	1	0.4666	0.1972

MONO: Monomorphic locus due to non polymorphism or missing data.

Supplementary Table 7. *C. callipygus* expected heterozygosity and Hardy-Weinberg Equilibrium P-values.

		He		Hardy-Weinberg Equilibrium test										
	n	**Mean**	**SD**	**BM2113**	**INRA40**	**BM1225**	**MM12**	**BM121**	**BM848**	**SR12**	**BM143**	**INRA05**	**BM1862**	**BM864**
HAO (32)	6	0.6157	0.2433	1	0.7131	0.04772	1	1	1	0.1097	MONO	1	0.08924	0.1998
IV_East (28)	11	0.6965	0.2091	0.01595	1	0.10783	0.1977	1	0.763	0.0027	MONO	MONO	MONO	MONO
IV_West (27)	7	0.7411	0.1889	0.00781	0.1643	0.07655	0.2735	0.1299	0.1037	0.0214	MONO	MONO	MONO	MONO
LA (29)	5	0.5328	0.2966	0.12652	0.3144	1	1	0.0467	1	0.0842	MONO	MONO	MONO	MONO
LBK (37)	11	0.5826	0.2575	0.01689	1	0.40579	1	0.2774	0.7248	0.0094	MONO	MONO	MONO	MONO
LO_North (23)	6	0.6049	0.1948	0.42957	0.3113	1	1	0.3995	1	0.0473	0.06569	1	0.42713	MONO
LO_South (24)	79	0.6512	0.2465	0.9007	0.0173	0.20551	0.049	0.0685	0.5381	<10-5	0.07308	0.88972	1	<10-5
MFCH (33)	4	0.6929	0.275	0.30332	0.6552	0.31241	MONO	0.4275	1	MONO	MONO	MONO	MONO	MONO
MKB (19)	7	0.6238	0.3177	0.49561	1	0.25378	1	0.0769	0.1603	0.0027	MONO	MONO	MONO	MONO
ND/CO (38)	14	0.5974	0.269	0.27434	0.4137	0.25393	1	0.4358	0.3552	0.0022	MONO	MONO	MONO	MONO
OIV_North (25)	12	0.6382	0.2308	0.1094	0.132	0.86418	1	0.3427	0.2435	0.0003	0.33208	1	1	0.06702
OIV_South (26)	17	0.5595	0.2874	0.52994	0.959	0.00928	1	0.7381	0.8098	0.0002	0.1971	1	1	1
OLO_North (30)	18	0.6285	0.1856	0.87832	0.9821	0.82085	1	0.1037	0.0901	0.0006	1	0.51605	MONO	0.51023
OLO_South (31)	18	0.569	0.2617	0.99706	0.0082	0.22308	1	0.3781	0.6112	0.0002	0.08477	0.19385	1	0.01578
WA (34)	4	0.6629	0.1845	0.65796	1	1	1	1	1	0.2003	MONO	MONO	MONO	MONO

MONO: Monomorphic locus due to non polymorphism or missing data.

Supplementary Table 8. *C. callipygus* summary statistics of Linkage Disequilibrium test

	n	BM2113	INRA40	BM1225	MM12	BM121	BM848	SR12	BM143	INRA05	BM1862	BM864	P-value
IV_West (27)	7	X	X										0.0003
LO_North (23)	6							X	X				0.0001
OIV_North (25)	12							X		X			0.0004
OLO_North (30)	18					X				X			0.0006
OLO_South (31)	18		X					X					0.00005

Supplementary Table 9. *P. monticola* expected heterozygosity and Hardy-Weinberg Equilibrium P-values.

		He		**Hardy-Weinberg Equilibrium test**										
	n	**Mean**	**SD**	**BM2113**	**INRA40**	**BM1225**	**MM12**	**BM121**	**BM848**	**SR12**	**BM143**	**INRA05**	**BM1862**	**BM864**
CPO (14)	18	0.788	0.1244	0.84977	0.7076	0.78739	0.0035	0.38767	0.2936	0.0066	MONO	MONO	1	MONO
Deng (10)	5	0.8492	0.0797	1	1	1	1	1	0.6593	1	MONO	MONO	1	MONO
Ebo (6)	10	0.872	0.0654	0.14647	0.3383	0.33494	0.6614	0.1995	0.039	0.2827	MONO	MONO	MONO	MONO
Edéa (5)	8	0.7808	0.1293	1	0.9098	0.05379	0.0254	0.6567	1	1	MONO	MONO	0.26959	MONO
GA (35)	6	0.7669	0.1085	0.5946	0.6581	0.58398	0.2536	0.0708	0.6548	0.3576	MONO	MONO	MONO	MONO
HAO (32)	4	0.7974	0.1727	0.46381	1	0.42525	1	0.19916	0.3126	0.1274	1	MONO:	1	MONO:
IV_East (28)	7	0.8485	0.149	1	0.5055	1	1	0.12722	0.1442	0.1138	MONO	MONO	MONO	MONO
IV_West (27)	7	0.7888	0.1769	1	0.3982	0.1649	0.7055	0.00996	0.0547	0.3989	MONO	MONO	MONO	MONO
LA (29)	5	0.822	0.062	1	1	1	1	0.08534	0.2372	1	MONO	MONO	MONO	MONO
LO_North (23)	5	0.6699	0.148	0.6922	0.1861	1	1	0.20061	1	1	MONO	MONO	1	MONO
LO_South (24)	28	0.7797	0.137	0.48795	0.2223	0.2461	1	0.00005	0.295	0.1304	MONO	0.33334	0.93602	0.08453
MCR (17)	8	0.8257	0.1256	1	0.1342	0.7821	0.491	0.20299	0.0508	0.0455	MONO	MONO	MONO	MONO
MFCH (33)	9	0.7794	0.1307	0.01617	0.038	0.22498	0.496	0.70273	0.5757	0.03	MONO	MONO	MONO	MONO
MM/MTA (16)	17	0.8397	0.0854	0.98487	0.7738	0.73833	0.1722	0.04498	0.4959	0.0458	MONO	MONO	MONO	MONO
MOO (21, 22)	6	0.8019	0.1615	0.39205	0.6139	0.61866	0.3934	0.46758	0.0607	0.0879	MONO	1	0.19396	1
ND/CO (38)	17	0.7783	0.1479	0.0638	0.0129	0.12299	0.2674	0.00784	0.0198	0.106	MONO	MONO	MONO	MONO
NGO (39)	55	0.8172	0.1078	0.01218	0.2089	0.07665	0.1189	<10-5	0.0037	0.001	1	0.33513	0.83113	0.04788
NIG (2)	6	0.8442	0.1373	1	0.0312	1	0.6359	0.19951	1	1	MONO	MONO	MONO	MONO
NKO (12)	5	0.8181	0.1719	0.01741	1	0.85	1	1	1	0.4656	MONO	MONO	1	MONO
OIV_North (25)	4	0.7057	0.1859	0.31343	0.7733	0.31686	1	1	0.4638	1	MONO:	1	1	0.20196
OLO_South (31)	6	0.7488	0.2233	0.83084	0.2867	0.41902	0.6242	0.00949	0.2979	0.0756	1	MONO	1	1
TAK (3)	6	0.7582	0.2335	0.71254	0.1686	1	1	0.02731	0.1538	0.3333	MONO	MONO	MONO	MONO

MONO: Monomorphic locus due to non polymorphism or missing data

Supplementary Table 10. *P. monticola* summary statistics of Linkage Disequilibrium test.

	n	BM2113	INRA40	BM1225	MM12	BM121	BM848	SR12	BM143	INRA05	BM1862	BM864	P-value
Ebo (6)	10	X	X										0.00055
LO_South (24)	28	X	X										0.0002
		X		X									<10-5

VITA

The author was born in Alhonha (near Lambaréné), Moyen Ogooué, Gabon. He obtained his Diplôme Universitaire d'Etudes Scientifiques, speciality Chemistry-Biology-Geology from the Université des Sciences et Techniques de Masuku, Franceville, Gabon in 2001. He obtained his Bachelor's degree in Organismal Biology and Master 1 in Environmental Biosciences both from the Université de Droit, d'Economie et Sciences d'Aix-Marseille III, Marseille, France in 2004 and 2005 respectively. He joined the University of New Orleans Biology graduate program to pursue a PhD in Conservation Biology and became a member of Dr Nicola Anthony research group in 2006.