Evolution of the African Tigerfish (Genus *Hydrocynus*):

Phylogeographic insights into drainage evolution

Sarah Goodier

Supervisors:
Dr Colleen O'Ryan (University of Cape Town)
Dr Woody Cotterill (University of Cape Town)
Prof Paul Skelton (South African Institute of Aquatic Biodiversity)
Prof Maarten de Wit (University of Cape Town)

Department of Molecular and Cellular Biology
Faculty of Science
University of Cape Town

In collaboration with:
Africa Earth Observatory Network (AEON)
University of Cape Town
and
South African Institute of Aquatic Biodiversity (SAIAB)
Rhodes University

Thesis presented for the degree of Master of Science

February 2010

A collaborative initiative between:

Plagiarism Declaration

"I know the meaning of plagiarism and declare that all of the work in the document, save for that which is properly acknowledged, is my own".

Signature: ____________________________

Date: ______________________________

Abstract

Tectonic changes exert controls over biodiversity evolution, which allows certain organisms to be used as biotic indicators of landscape history in any given area. Such species can potentially resolve questions about key events that changed landforms. Aquatic species, notably fish, can reveal details of drainage evolution, especially where their evolutionary relationships can be reconstructed using phylogeographic methods. In this study, the mitochondrial DNA sequence diversity of the characiform genus, *Hydrocynus*, which is widespread across tropical Africa, is reconstructed using a phylogenetic framework and divergences are dated using the *cytochrome b* (*cyt b*) region. The constructed phylogeny reveals ten unique evolutionary lineages, more than the five species currently recognised based on morphology. *H. tanzaniae* is found to be sister to a group consisting of several novel lineages (Groups B, C and D) and *H. vittatus* (collectively called the Vittatus complex). The dated tree reveals that the divergences within the Vittatus complex occurred through the Plio-Pleistocene. These events reflect changes to drainage systems across south-central Africa. More detailed phylogeographic analyses of the Vittatus complex using the mtDNA control region (CR) genetic marker reveals finer-scaled details of these geomorphological events. The CR analyses recover a phylogeny congruent with that obtained using the *cyt b* sequences, which confirms the strong geographical structuring observed within the Vittatus complex. This result indicates that *H. vittatus* diverged as a result of the severing of drainage connections between the Zambian Congo system and the Kafue River (where *H. vittatus* subsequently died out), which founded the Upper Zambezi *H. vittatus* population. This population, in turn, likely seeded all other populations of *H. vittatus* to the north (Congo) and east (Middle and Lower Zambezi; south coastal drainages). Further sampling and analysis, with tighter dating constraints, need to be conducted to expand on the findings of this study.

Acknowledgements

This project would have been impossible without many individuals and organisations who have given help and assistance. I would like to thank my main funders: the Deutscher Akademischer Austauschdienst (DAAD) and the John Ellerman Foundation for their generous support. I am also grateful for the additional aid I received in the form of the Harry Crosley Foundation Scholarship, the Twamley Bursary, the KW Johnston Bequest and the Whitehead Scientific Travel Grant.

Thanks also go to all those who helped with obtaining samples and provided accommodation and all kinds of practical assistance, without which this project would not have been possible: Vanessa Anastassiou; Roger Bills; Dr Jacqui Bishop; Sven Bourquin; Ian and Emma Bruce-Miller; Ryan Clark; Keith Clover and Tourette Fishing Safaris; Guy and Luke Coulson; Dr Natalie Coutts; Patrick Danckwerts; Mike de Wit and BRC-Diamond; Mike and Jan Fisher; Robin Hartle and the guides, staff at Nxamaseri, Guma and Drotsky's Camps; Michel Hasson; Anthony Kaschula; Sally Kelly; Bastien Linol; Paul Lombard; John Minshull; Andy Moore; Stan Nabozny; Dirk Neumann; Poogendri Reddy; Dr Martin Reichard; Dr Hugh Roberts; Francios Roux; Jihad Salhab; Dr Bob Schelly; Ernst Schwartz; Ulrich Schweilen; Dr Melanie Stiassny; Denis Tweddle; Kevin Wallace; Craig Zytkow and Ndole Bay Lodge; as well as the National Geographical Society and Ndola Fisheries.

Finally, several individuals deserve a special thanks:

- My supervisors, Dr Colleen O'Ryan, Dr Woody Cotterill, Professor Paul Skelton and Professor Maarten de Wit, for providing me with the opportunity and means to work on this exciting project as well as for their encouragement, support and my first international travel experiences.
- All my friends and my colleagues in Lab 405 and AEON, who were a constant source of support, inspiration, encouragement and saved my sanity by making me take breaks!
- My family and Tyrel, for encouraging me, listening to and supporting me when I needed it and being a sounding board for ideas.

Table of Contents

page

Appendices

List of Figures

page

List of Tables

Glossary

Bioturbation	The reworking and transport to the surface of sediments by termites and other burrowing animals.
Capture elbow	A sharp change in direction at the point of capture of a river.
Channel diversion	The redirection of a river into a neighbouring drainage system by breaching the drainage divide, possibly as a result of tectonics.
Epeirogeny	Uplift of a broad region.
Haplotypes	Unique sequences.
High Africa	Surface elevations of southern and eastern Africa over 1000m (rugged in parts).
Introgression	Hybridization between species.
Lithosphere	The outer solid part of the earth consisting of the crust and outer mantle.
Low Africa	Surface elevations of central and west Africa less than 500m (smoother topography).
Phylogeography	The study of the geographical distribution of genealogical lineages.
River beheading	The capture of parts of the headwaters of a neighbouring catchment that leaves little evidence of the abandoned channels.
River capture	Headwater erosion causing the diversion of one river into another resulting in a capture elbow.
Topotypical	Material obtained from the type locality.
Warp	An upward or downward flexing of the earth's crust.

List of Abbreviations and Symbols

Abbreviations

BEAST	Bayesian Evolutionary Analysis by Sampling Trees
bp	base pairs
BSA	Bovine Serum Albumin
CI	confidence interval
CR	control region
cyt b	*cytochrome b*
dNTP	deoxynucleotide triphosphate
EARS	East African Rift System
G	Gamma
gsi	genealogical sorting index
GTR	General Time Reversible
Hap	haplotype
Hd	haplotype diversity
HKY	Hasegawa, Kishino and Yano
hLRT	hierarchical likelihood ratio tests
indel	insertion or deletion
ka	thousand years ago
kb	kilobases
LGM	Last Glacial Maximum
Ma	million years ago
MCMC	Markov chain Monte Carlo
$MgCl_2$	magnesium chloride
min	minutes
mtDNA	mitochondrial DNA
numts	nuclear copies of mtDNA
OKZ	Ovamboland-Kalahari-Zambezi
Pers. comm.	Personal communication
rpm	revolutions per minute
RpS7	Ribosomal protein S7
SDS	sodium dodecyl sulfate

sec	seconds
SRTM	Shuttle Radar Topography Mission
s.s.	*sensu stricto*
TBE	tris-borate-EDTA
TBR	Tree-Bisection–Reconnection
Ti	transitions
Tv	transversions
UPGMA	Unweighted Pair Group Method with Arithmetic Mean

Symbols

Ne	effective population size
ΦST	fixation index
α	Gamma shape parameter
μL	microliters
μM	micromolar
mM	millimolar
mg/μL	milligrams per microliter
μ	mutation rate
U	units

Chapter 1

Introduction

1.1 General Introduction

Processes that alter landscapes are important in the history of the biodiversity in an area (Cotterill, 2004). Therefore, using biological indicators of landscape change can aid in resolving questions about geomorphological events and palaeogeography. In river systems, indicator species of fish can be used. In this study, the genetic history of an indicator genus of fish, *Hydrocynus*, which occurs in drainage systems[1] across most of Africa, is looked at through a phylogenetic framework and the major divergence events within this genus are dated. Genetic data from several closely related *Hydrocynus* lineages including *H. vittatus* (collectively named the Vittatus complex in this study) are analysed using phylogeographic methods and what this reveals about the evolution of drainage systems in south-central Africa[2] is discussed.

The various sections of this introduction explore the main relevant themes of the study. First, a description of the study area (section 1.2) is provided, covering the drainage systems, both ancient and modern. A review of the literature concerning the indicator of drainage change used in this study, namely *Hydrocynus* and particularly *H. vittatus*, is also presented (section 1.3). This study comprises two components that report new genetic data, each of which is the focus of a separate section. The first presents the phylogenetic relationships of *Hydrocynus* using mitochondrial DNA (mtDNA) evidence in the form of the *cytochrome b* (*cyt b*) marker (section 1.4). The second focuses mainly on one

---

[1] For the purpose of this study, a drainage system is defined as a major river with its tributaries and lakes, delimited by watersheds.

[2] South-central Africa is defined as the area including and south of the Congo drainage system.

species, *H. vittatus*, and the lineages most closely related to it using a phylogeographic analyses of control region (CR) marker data (section 1.5). Both of these main sections examine how genetics in these fish relates to the drainage system changes of Africa discussed in section 1.2. The second component also explores how the evolution of *H. vittatus* and closely related lineages relates to landscape evolution across south-central Africa.

1.2 Physiographic Description of Study Area

Africa's modern day outline has been in place since the early Cretaceous (Roberts, 1975) and the continent as a whole has been relatively stable since the breakup of Gondwana, approximately 130 to 120 Ma (Reeves & de Wit, 2000) (see Figure 1.1 for a geological time scale). In the description of the study area of this project, the topography of Africa and its major rift system will be briefly described in order to provide a general context (section 1.2.1). The relevant drainage systems, both their modern topology (section 1.2.2) and what is known from the literature of their ancient topologies (section 1.2.3), is also summarised. This review reflects the state of knowledge of the geomorphology of the African landscape across scales of evolutionary space and time to the present day, providing a geomorphological framework within which to examine the genetic data gathered in this study.

1.2.1 Bimodal Africa and the East African Rift System

Africa's topography is distinctly bimodal and has a complex history. It is the result of recurrent episodes of uplift, erosion and tectonics. Surface elevations of southern and eastern Africa average over 1000 m and are rugged in parts (High Africa) while central and west Africa are smoother with elevations less than 500m (Low Africa) (Roberts, 1975; de Wit, 2007) (see Figure 1.2). The geophysical reconstruction of Doucouré and de Wit (2003) concluded that this topography is a feature from at least the Mesozoic. The Kalahari plateau of southern and central Africa, a part of High Africa, is the main focus of the current study. This elevated area of low topographic relief extends from the Cape Fold Belt northwards to the Congo basin. It was formed during the Kalahari epeirogeny, which refers to hundreds of millions of years of uplift thought to be

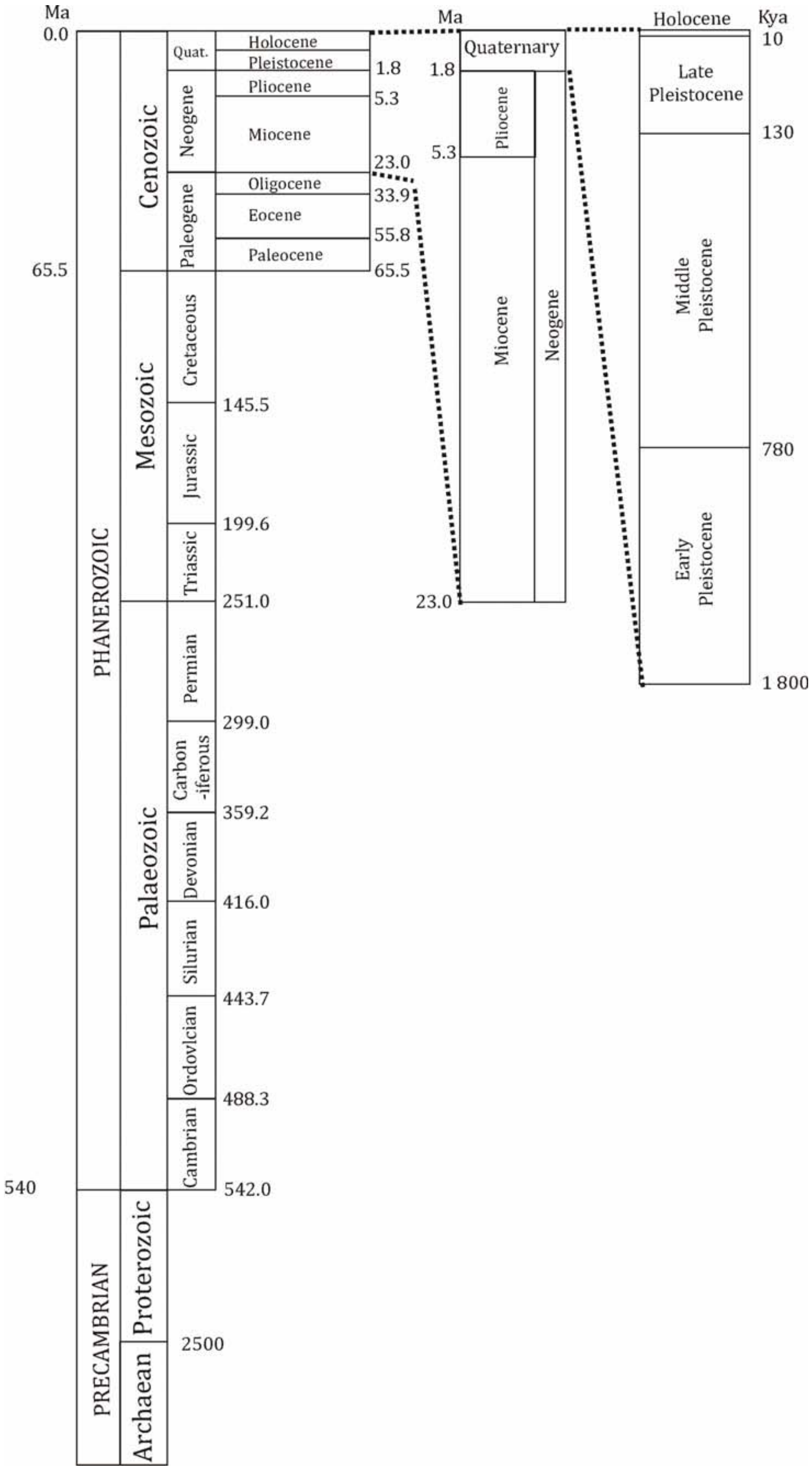


Figure 1.1: The geological time scale
Based on the international stratigraphic chart 2009 (Ogg, 2009).

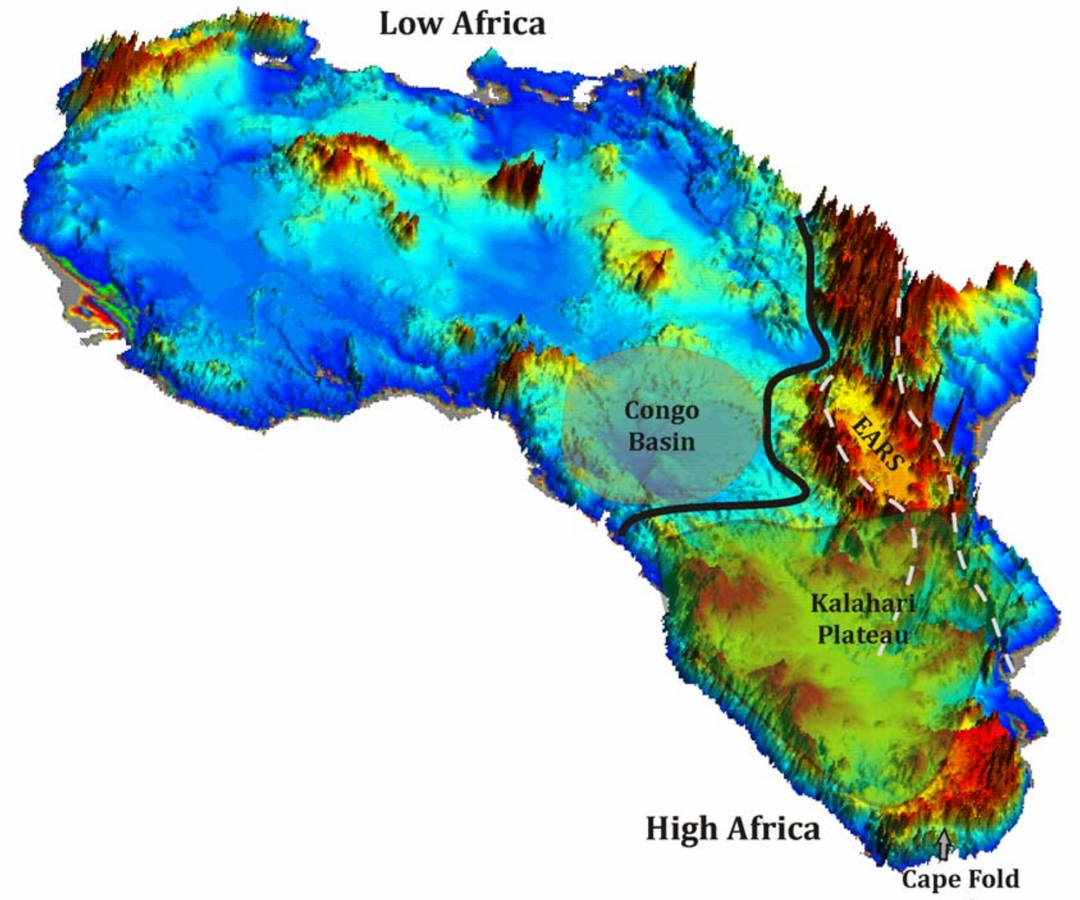


Figure 1.2: A three-dimensional perspective of bimodal Africa
Modified from Doucouré and de Wit (2003).
Highest and lowest elevations are indicated in red and blue, respectively. The thick black line designates the divide between High and Low Africa. The rifts of the EARS are indicated by light grey dashed lines.

caused by mantle plumes pushing the lithosphere upwards, erosion (Goudie, 2005; de Wit, 2007) and plate movements (Moore *et al.*, 2009a).

This uplift caused the rejuvenation of several rivers around the plateau, particularly on the coast (Allanson *et al.*, 1990). Today, this plateau contains several major drainage basins and flanking coastal plains. The rivers that drain the High African plateau often have low gradients for much of their course, whereas the coastal rivers are steep, eroding into the edges of the plateau. Through the Cenozoic, these rapidly eroding rivers have captured several of the rivers on the plateau, diverting them to the coastal systems (Moore & Larkin, 2001). The Congo basin, as its name suggests, is a part of Low Africa. This oval depression, which contains the Congo River, is between 300 and 500 m in elevation and has its origin in at least the Palaeozoic (Runge, 2007).

The East African Rift System (EARS), which originated in the Eocene/Oligocene (Burke & Gunnell, 2008), is an important geological feature of Africa (Figure 1.2). This rift system acts as a catchment for water, containing most of the Great Lakes of East Africa (e.g. Lakes Tanganyika and Malawi) (Lévêque, 1997). The EARS consists of two main branches: the Eastern branch extends north from the Mozambique coastal plain through the Shire Valley and Lake Malawi to Lake Turkana and through the highlands of Ethiopia and across to the Red Sea; the Western branch starts near Lake Malawi and extends north through Lakes Tanganyika, Kivu, Edward and Albert. The south west extension of this rift extends to the Okavango drainage system (Gumbricht *et al*., 2001; Cotterill, 2006). The rift system has undergone repeated and ongoing faulting and volcanic activity, initiated in the late Oligocene. These rifts are important biogeographical barriers between southern and eastern Africa, constraining the distribution of many taxa, including fish (Roberts, 1975) and terrestrial vertebrates (Grubb *et al.,* 1999; Cotterill, 2003).

1.2.2 Modern Drainage Systems

Africa has experienced periods of tectonic stability interspersed by uplift along axes in the Miocene and Pliocene, which has affected rivers across the landscape (Dollar, 1998; Goudie, 2005) and, as a result, influenced fish evolution (Skelton, 1994). It is this complex history of epeirogeny, along with Africa's geology, that is the main factor since the breakup of Gondwana which is cited as resulting in the development of the continent's modern drainage systems (Lévêque, 1997; Dollar, 1998). Level changes within systems created by uplift have resulted in the formation of waterfalls and rapids, isolating certain fish faunal elements, and causing an increase in erosion rates and the rejuvenation of certain rivers (Lévêque, 1997).

As the main geographical focus of this study is south-central Africa, the drainage systems of this area, including possible barriers to fish dispersal, will be discussed in some detail below while the other relevant systems to the north, west and east will receive only brief treatment. This section has been divided into several freshwater icthyofaunal provinces (broadly as per Roberts (1975)) since aquatic biodiversity can be separated into several regions based on species' distributions and the similarity of the fauna between different rivers (Abell *et al.*, 2008). The principal freshwater icthyofaunal provinces relevant to this study are the Congo and Zambezi. The Nilo-Sudan and Eastern and Coastal icthyofaunal provinces are of more peripheral relevance (Figure 1.3).

Congo Icthyofaunal Province

The Congo icthyofaunal province (Province II, Figure 1.3) encompasses the Congo and Zambian Congo systems as well as Lake Tanganyika. The Congo River, which runs through a large basin in central Africa, is the largest drainage system in Africa and is also the most species rich (Roberts, 1975). The Lualaba River, which represents the Upper Congo and is characterised by a series of prominent rapids over its last 150 km, flows north from its headwaters adjacent to the Zambian border close to the city of Kisingani. The Central Congo portion of the river continues from the Lualaba and swings to the west in an arc towards the coast. This section of the river ends with a series of rapids over 350 km. The Lower Congo River drops 275m to the narrow coastal plain. A major barrier on the Lualaba postulated to separate the Upper and Lower Congo fauna is a set of gorges and rapids known as the Portes d'Enfer or Gates of Hell (Malaisse, 1997). The Kasai River is one of the main north-flowing tributaries of the Congo and is equivalent to the mainstem in size (Thieme *et al.*, 2005). The Kwango is a major tributary of the Kasai (Thieme *et al.*, 2005). The Lower Congo River flows from below Pool Malebo (Stanley Pool) through a series of rapids to the coast. There are several large waterfalls and rapids on these tributaries that could prevent fish movement (T. J. Flügel, unpublished PhD thesis results (pers. comm.)).

Lake Tanganyika is thought to be approximately between 9 and 12 Million years old (Sturmbauer *et al.*, 2001). It is connected to the Upper Congo via the Lukuga River. The Lukuga flows out of the lake and the high waterfalls on the river are postulated to be a barrier to upstream fish dispersal (Katongo *et al.*, 2005). Lake Tanganyika reaches a depth of 1470m and oxygen levels below approximately 200m are too low to support aerobic life, with the result that most of the fauna and flora are confined within the upper reaches of the lake. The outflow from

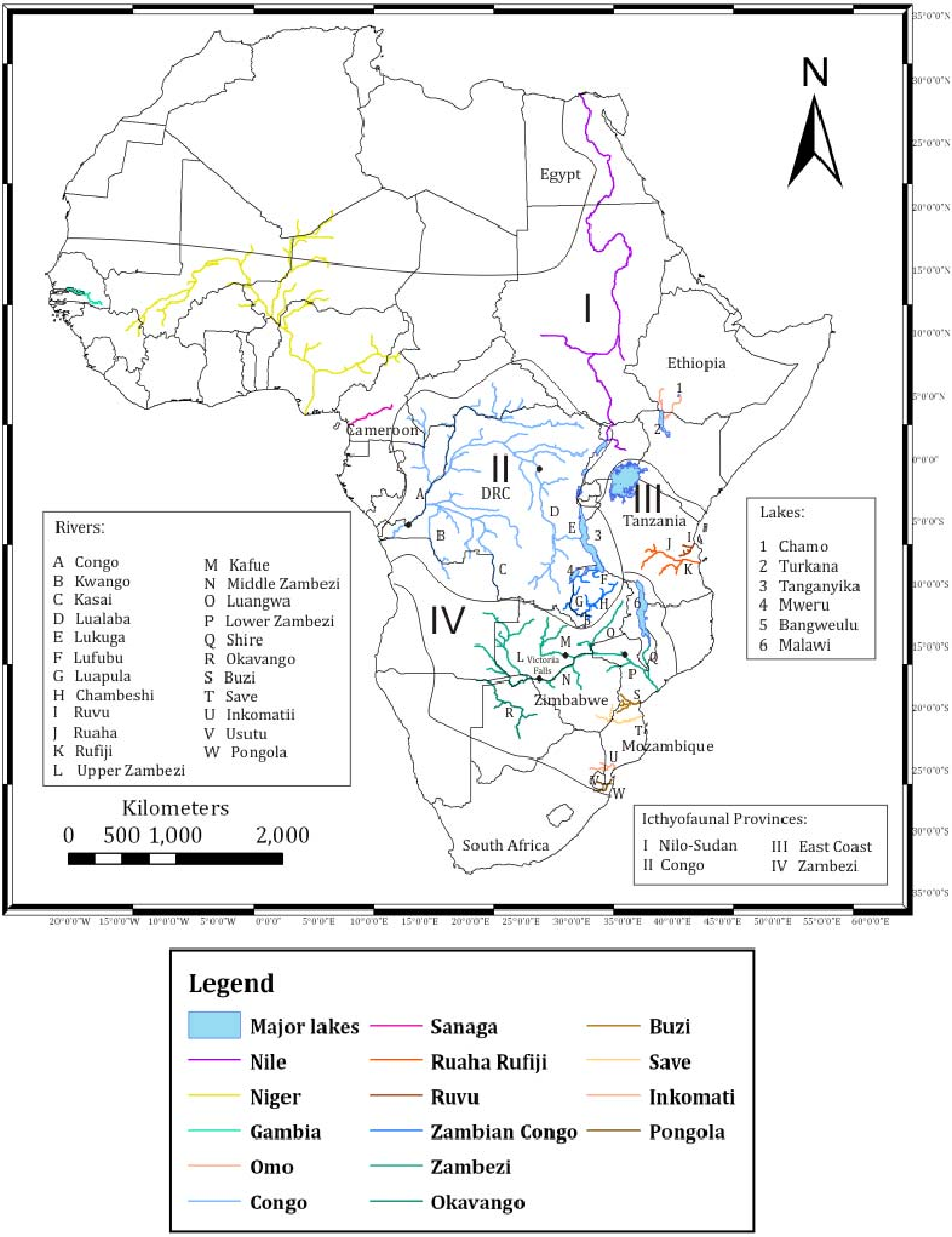


Figure 1.3: Map of the relevant main river channels showing Africa's four main freshwater ichthyofaunal provinces dealt with in this study

This map was constructed using ArcMap 9.3 from the HydroSHEDS digital river database (Lehner *et al.*, 2008). Provinces are labelled with Roman numerals. Main rivers and lakes discussed in the text are labelled with letters and numbers, respectively. Black dots represent selected barriers to fish movement discussed in the text. Political boundaries and country labels are included for geographical context.

Lake Tanganyika to the Congo is weak and in the past it has been intermittent (McClanahan & Young, 1996). The Lufubu River is the prominent west bank tributary of Lake Tanganyika at its southern end.

The Bangweulu basin is suggested to have been formed via crustal warping (Dixey, 1944), which created an inland drainage basin. The Chambeshi River feeds Lake Bangweulu (Debenham, 1947), which is connected to Lake Mweru via the Luapula River. The Luapula has two sets of falls and rapids which are thought to restrict fish movement between the two lakes in the dry season but when the river is in full flood, these falls submerge partially and fish movement is thought to occur (Bell-Cross, 1965a). Although these two lakes appear to share many species (Bell-Cross, 1965a), the fish biodiversity of the area is still poorly known overall (Jackson, 1986 & 2000).

A rich and diverse fish fauna has evolved in the Congo basin. However, this drainage system has not been studied in any great detail (Lévêque, 1997), as indicated by ongoing discoveries of new species and genera (e.g. Banister & Bailey, 1979; Reid, 1996; Stiassny & Schliewen, 2007).

Zambezi Icthyofaunal Province

The Zambezi province (Province IV, Figure 1.3) is composed of the Zambezi and Okavango systems as well as those river systems down the southern East Coast reaching as far as the Mkuze River (the drainage system to the south of the Pongola). The Okavango is the only extant river in south-central Africa terminating in an inland delta. Waters from the Okavango system intermittently overflow into the Upper Zambezi system via the Selinda spillway during heavy rains and the fauna of both are highly similar (Bell-Cross, 1965b & 1972; Roberts, 1975).

The Zambezi system is the largest eastward-flowing river system in the study area. It is divided into three sections by physical barriers (Roberts, 1975; Jackson, 1986; Skelton, 1994; Moore & Larkin, 2001). The Upper Zambezi extends from the headwaters to the Victoria Falls, initially flowing south-east before changing to an easterly course that has been postulated to be a capture elbow[3] where the Middle Zambezi beheaded the Upper Zambezi (Wellington 1955). Along the Upper Zambezi, there are two stretches of rapids. The Ngonye Falls is the only waterfall along this river that serves as a barrier to the movement of fish but only in the dry season and no difference in fish distribution above and below these falls has been observed (Bell-Cross, 1972). The Victoria Falls have been postulated to be a major barrier to upstream dispersal of fish fauna from the Middle Zambezi into the Upper Zambezi (Roberts, 1975). However, there is a recurring debate as to whether the falls prevent downstream dispersal of Upper Zambezi fish as well (e.g. Bell-Cross, 1965a & b; Balon, 1974; Jubb, 1977) (see section 1.3.3). The Upper Zambezi has a more diverse fish fauna compared to the Middle and Lower Zambezi (Balon, 1974; Jackson, 1986).

The Middle Zambezi extends from the Victoria Falls to the Cabora Bassa gorge and contains two large dams, Lake Kariba and Cabora Bassa. These dams have provided a new environment for aquatic fauna within this stretch of the Zambezi, which has caused some fish to decline (e.g. *Opsaridium zambezense*) while other species have thrived (e.g. *Oreochromis* species) (Jackson, 1986). The Kafue and Luangwa Rivers are the major north bank tributaries of the Middle Zambezi. These rivers also undergo sharp changes in topology from southerly and south-west to easterly and south-east, respectively. Both changes have been interpreted as capture elbows (Thomas & Shaw, 1991; Moore & Larkin, 2001). In the Kafue Gorge the Chasunta Falls prevent movement of Middle Zambezi fish upstream into the Kafue (Bell-Cross, 1972).

The Lower Zambezi starts at the Cabora Bassa gorge and ends in a delta flowing into the Indian Ocean. This section of the river flows over the Mozambique coastal plain and its major tributary is the Shire River which flows from Lake Malawi. In the Middle Shire, the Kapachira Rapids provide a biogeographical barrier to upstream dispersal by freshwater taxa. The Middle and Lower

---

[3] A capture elbow is defined as a diversion of one river into another resulting in a sharp change in direction at the point of capture.

Zambezi, including the Shire below the Kapachira rapids are classified as one biogeographical unit (Tweddle *et al*., 1979; Skelton, 1994).

The Mozambique coastal plain is broadest in northern Mozambique and narrows substantially to the south (Skelton, 1994). The Buzi, Save, Inkomati and Pongola Rivers all have their headwaters in the eastern escarpment highlands, including the Chimanimani and Drakensburg mountain areas, from which they flow across this coastal plain into the Indian Ocean. The Save River initially flows south before turning east across the Mozambique coastal plain (Wellington, 1955).

Nilo-Sudan Icthyofaunal Province

The Nilo-Sudan province (Province I, Figure 1.3) extends across most of North Africa, encompassing the major drainage systems of the Nile, the Niger, the Gambia, the Omo and the Sanaga. The Nile, the longest river in Africa, has its headwaters to the north-east of Lake Tanganyika in the Lake Victoria and Lake Albert catchments and ends in a delta flowing into the Mediterranean Sea. Its main tributaries are the White and Blue Nile. The present topology of the Nile dates from the Pleistocene (Potter and Hamblin, 2006). Roberts (1975) describes Lake Chamo as an isolated lake and only receives flow from nearby drainages in flood. However, it is thought that it previously had a connection with Lake Turkana and the Omo system (Isaac & Behrensmeyer, 1997; Woldegabriel *et al*., 2000).

The Niger River, the longest river in west Africa, drains a large section of north-west Africa. It has its origin in the Guinea highlands (Lévêque, 1997) and flows out into the Atlantic Ocean through a delta. Roberts (1975) hypothesises that the Niger River was a source from which the Gambia and Senegal Rivers were recolonised after nearly drying out in the last dry climatic period (approximately 27 - 12 ka) (Lévêque, 1997).

The Gambia River has its origin on the northern Guinea plateau. It flows through southern Senegal and the Gambia into the Atlantic Ocean (Albaret *et al*., 2004). The fauna in the upper Gambia section are distinct from those in the middle and lower sections (Daget, 1960). The Sanaga River flows through a mountainous area in Cameroon into the Atlantic Ocean and although it is the largest drainage system in Cameroon, its total catchment area is small relative to the larger systems in Africa (Shahin, 2002).

East Coastal Icthyofaunal Province

The East Coastal province (Province III, Figure 1.3) encompasses coastal rivers from Kenya to Mozambique north of the Zambezi and includes the Ruvu and the Ruaha-Rufiji systems, which flow into the Indian Ocean. The Ruaha-Rufiji system is the largest drainage basin in Tanzania (Temple & Sundborg, 1972). The Ruaha is a major tributary of the Rufiji, which flows into the Indian Ocean. The Ruvu flows down from the escarpment onto the coastal plain. The fish fauna in this eco-region is relatively poor and Roberts (1975) has suggested that this is due to the drying out of these systems during the Pleistocene interpluvials and that current fish fauna therefore represent colonisation events in the past 12 000 years. This region has links to the Zambezi fauna (Skelton, 1994; Lévêque 1997).

1.2.3 Drainage Rearrangements

The drainage systems of south-central Africa have been investigated in several geological and geomorphological studies (e.g. Dixey, 1945; Dollar, 1998; Moore & Larkin, 2001). Since the break-up of Gondwana in the mid-Jurassic, these systems have experienced several major rearrangements, which are the culminations of the interactions of complex geological processes, glaciation and tectonics (Bishop, 1995), as well as the influence of palaeo-climate (Partridge, 1997). However, much is still unresolved concerning the drainage evolution and geomorphology of south-central Africa and its underlying causes. There is a lack of precision of dating events in the Quaternary, approximately 1.8 to 0 Million years ago (Ma), and knowledge of drainage evolution in the Neogene is poor (Cotterill, 2006). This is due to a combination of factors: the limits of traditional geological absolute dating techniques, uncertainties related to interpreting the available data, incompatible conclusions from evidence and, above all, the lack of availability of geological material and the limited number of studies conducted across the region (Thomas & Shaw, 2002; Watchman & Twidale, 2002).

In addition, intense, relatively recent erosion has taken place over large areas of the uplifted Kalahari plateau and the reworking of dunes in the area during the Quaternary has reset the dating signals, making the dating of landforms in this region difficult (Thomas & Shaw, 2002; Bateman *et al.*, 2003) and introducing uncertainty into drainage reconstructions (Bishop, 1995). Sediment dating is an important tool in reconstructing a timeline of geomorphological events affecting drainage but bioturbation can have a major effect on sedimentary sequences leading to incorrect dates of sediments (Johnson, 2002). .

Geologically-based drainage hypotheses relevant to this study are presented below in the context of the geological time scale period over which they appear to have occurred (Table 1.1; Figure 1.4). Where supporting biogeographical information is known, it is included only to compare with such geological hypotheses since it is circular to infer drainage history based on the distribution of fish species only to use this drainage hypothesis to explicate fish distributions (Bishop, 1995). A brief discussion of the effect of palaeo-climate on African fish will be included. Bell-Cross (1965a) has proposed separate eastern and western drainage systems with characteristic fish faunas in western, interior and eastern drainage basins at the start of the Tertiary. This is supported by Skelton (1994) and is used in this study as the starting point in this study from which changes are presented.

Table 1.1: Summary table of the drainage system changes from the Jurassic to the end of the Late Pleistocene

	Time Period		
River	**Jurassic -Early Cretaceous** (±200-100 Ma) (see Figure 1.4a)	**Late Cretaceous** (±100-65 Ma)	**Paleocene** (±65-55 Ma) (see Figure 1.4b)
Chambeshi	Headwaters of the Kafue[1] OR Tributary of the Congo-Rufiji system[2]	-	-
Kafue	Flowed into Upper Zambezi[1,3,4]	-	-
Luangwa	Flowed into Upper Zambezi[1,3,4] OR Tributary of the Congo-Rufiji system[2]	-	-
Middle Zambezi	Flowed into Upper Zambezi[1]	-	-
Upper Zambezi	Flowed into the Limpopo[1,3,4] *	-	Link with Limpopo severed by uplift along Ovamboland-Kalahari-Zambezi (OKZ) Axis[9], creating major inland drainage system[1, 4,10]
Okavango	Flowed into the Limpopo[1,5,6]	-	(As for the Okavango - see above)
Limpopo	-	-	-
Save	-	Incising into the coastal margin[1]	Underwent headwater erosion[1]
Lower Zambezi	Established after Gondwana breakup[1,4]; head water erosion initiated[7]	Continued incising into the coastal margin[1]	Underwent headwater erosion[1]
Shire	Established after Gondwana breakup[1,4]	-	-
Congo (basin)	Large lake[8]; possible link to the Rufiji system[2]	-	Uplift around Congo basin[11]
Rufiji	Drained the Congo into the Indian Ocean[2]	-	-
Luapula	-	-	-

- : No major changes are proposed to have occurred;
* : Similar fish fauna occur in the Limpopo & Zambezi systems today (Gaigher & Pott, 1973; Skelton, 1994)

[1] Moore & Larkin (2001); [2] Stankiewicz & de Wit (2006); [3] Dixey (1945); [4] Thomas & Shaw (1991); [5] Du Toit (1927); [6] Du Toit (1933); [7] Moore *et al.* (2009b); [8] Peters & O'Brien (2001); [9] Moore (1999); [10] Grove (1969); [11] Giresse (2005); [12] Burke (1996); [13] Goudie (2005); [14] Bell-Cross (1965b); [15] Sturmbauer (2001); [16] Wellington (1955); [17] Derricourt (1976); [18] Dixey (1943); [19] Cotterill (2006); [20] Moore & Cotterill (2009); [21] Cotterill (2004); [22] McClanahan & Young (1996)

Table 1.1: continued

	Time Period		
River	**Eocene – Oligocene** (±55-23 Ma) (Figure 1.5a)	**Miocene - Early Pleistocene** (±23-1.8 Ma) (Figure 1.5b)	**Middle - Late Pleistocene** (±1.8 Ma – 10 ka) (Figure 1.5c)
Chambeshi	-	Captured by Luapula, creating Lake Bangweulu[1,18]	-
Kafue	-	-	Flowed into Okavango[19,20]; recently captured by Middle Zambezi[21]; connections with Chambeshi severed[21]
Luangwa	Beheaded by Lower Zambezi[1]	-	-
Middle Zambezi	-	Captured Upper Zambezi[1,7,15]	-
Upper Zambezi	-	Captured by Middle Zambezi[1,7,15]	Intermittently flowed into Okavango & captured by the Middle Zambezi [1,19,20]
Okavango	-	-	-
Limpopo	-	-	-
Save	-	-	-
Lower Zambezi	Beheaded the Luangwa[1]	-	-
Shire	-	-	-
Congo (basin)	Link with Rufiji severed[2]; captured by stream draining into the Atlantic[12,13] ▲	Watershed with Upper Zambezi moved south[14]; Lake Tanganyika formed[15] •	Lukuga River becomes an outlet for Lake Tanganyika[22]
Rufiji	Link with Congo severed[2]	-	-
Luapula	-	Captured Chambeshi & modern Upper Luapula, Lake Mweru deepened[1,18] #	-

▲: Sedimentary evidence lend support to this (Lavier *et al.*, 2001)

•: Similar fish fauna occur in the Congo & Upper Zambezi (Skelton, 1994)

#: Similar fish fauna in occur in the Congo & Middle & Lower Zambezi (Skelton, 1994)

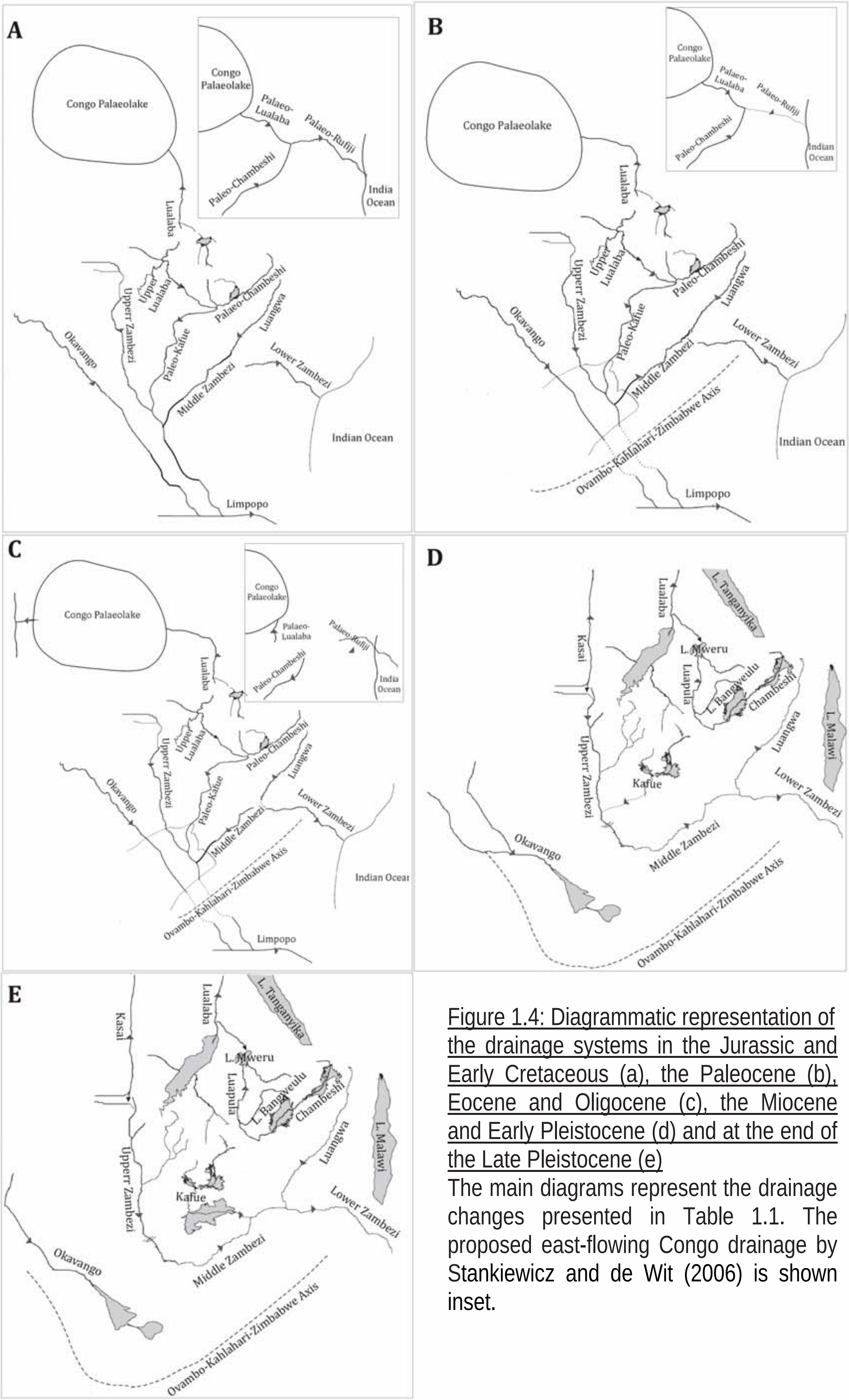


Figure 1.4: Diagrammatic representation of the drainage systems in the Jurassic and Early Cretaceous (a), the Paleocene (b), Eocene and Oligocene (c), the Miocene and Early Pleistocene (d) and at the end of the Late Pleistocene (e)
The main diagrams represent the drainage changes presented in Table 1.1. The proposed east-flowing Congo drainage by Stankiewicz and de Wit (2006) is shown inset.

The effect of palaeo-climates on African fish faunas and their biogeography is poorly understood. In fact, the palaeo-climatic conditions in Africa are debated with no clear consensus (Gasse, 2000; Boelhouwers & Meiklejohn, 2002; Thomas & Shaw, 2002) due to vast gaps with no primary palaeo-climatic data (e.g. the Congo and Zambezi basins) and differences in interpretation of the results gathered (Gasse, 2000). The recent ice ages caused colder and drier periods during the glacials which were interspersed with wetter interglacials (Hewitt, 2004). The most recent dry period, the Last Glacial Maximum (LGM), occurred between 27 and 12 ka and this is believed to have impacted on drainage systems and also caused lakes to dry up (Roberts, 1975). An example of this is the desiccation of Lake Malawi (Lyons *et al*., 2009). A more humid period is thought to have occurred approximately 12 to 8 ka and all lake levels were higher during that time (Roberts, 1975).

1.3. Indicator taxon: *Hydrocynus* and *H. vittatus*

As discussed above, dating of sediments is a tool to reconstructing a timeline of geomorphological events with confidence. Yet, whenever this is not possible or limited, other techniques should be tried to constrain when events occurred in these time periods. Using biological indicators of landscape change can help to resolve questions about geomorphological events because the same processes that alter the landscape are important in the history of the biodiversity living in the area (Cotterill, 2003 & 2004). Thus, research into the biogeographical and genetic patterns of extant species can give insights into the evolutionary history of a landscape. In this context, biogeographical data in the form of congruent distributions of divergent species can imply that a common cause has been significant in altering drainage topology in south-central Africa (e.g. Skelton, 1994; Cotterill, 2003).

Extending this argument, it is reasonable to postulate that signals of landscape changes have also been preserved in species' genomes. Where populations have been isolated over evolutionary time, genetic differences accumulate amongst separated populations (Bishop, 1995). These differences can reveal details of where and when the geomorphological events which altered landscape structure changed the drainage systems. The amount of genetic divergence between two, or more, isolated populations can, therefore, be expected to correlate with the period of geographical isolation (Bishop, 1995; Burridge *et al.*, 2007; Craw *et al.*, 2008). These dynamic links between landforms and biota can help to provide temporal resolution on such geomorphological events and shed light on their underlying causes (Cotterill & Goodier, 2008).

These links are particularly pronounced where the indicator species distribution is spatially limited by a dynamic drainage system that has been modified by geological processes such as tectonism (Avise, 2000; Burridge *et al.*, 2007), such as in the case of freshwater fish of any region (Bishop, 1995). Geological changes that cause events such as river capture, river reversal and the development of new drainage divides can cause the transfer of fish from one drainage system to another (i.e. dispersal) or cut off once connected systems (i.e. vicariance), leading to the accumulation of genetic differences between the fish contained within the separated systems (Bishop, 1995; Hurwood & Hughes, 1998; Burridge *et al.*, 2007). The interwoven relationship between the landscape and its aquatic fauna described above means that the evolutionary genetic history of fish populations can be used to infer relationships between different drainages (Bell-Cross, 1965a; Balon, 1974; Jubb, 1977; Avise, 2000; Burridge *et al.*, 2007) through the accumulation of population-specific genetic differences that represent their ancestral drainage systems (Bishop, 1995).

The genus *Hydrocynus* (Cuvier 1816) and specifically *H. vittatus* were selected for the current molecular study in order to examine relationships within and between drainage systems. These freshwater teleosts, commonly referred to as tigerfish, are members of the Order Characiformes (Lévêque, 1997; Calcagnotto *et al.*, 2005). This section gives an overview of the genus, whose genetic data will be examined in a phylogenetic framework (see section 1.4 and Chapter 2). This will be followed by information on *H. vittatus* specifically. The genetic data from this species will be examined in a phylogeographic context (see section 1.5 and Chapter 3).

1.3.1 The Genus *Hydrocynus*

Hydrocynus are dominant predators in African freshwater systems. Members of this genus are postulated to structure the fish fauna in African rivers by restricting the use of the main channels to larger fish and those that have morphological features that minimise the risk of predation, such as dorsal and pectoral spines (e.g. found in *Synodontis*) or a deep body (Jackson, 1961), features which make it more difficult for *Hydrocynus* to swallow the prey. All species of *Hydrocynus* are sought-after game fish, which are important in tourism (Gaigher, 1970; Kenmuir, 1973; Meintjies, 1999; Næsje *et al.*, 2001; Meintjies, 2005). In addition, *Hydrocynus* are an important component in subsistence and commercial fisheries, providing an essential source of dietary protein (Skelton, 2001).

There are currently five recognised species, based on morphological analysis: *H. forskahlii* (Cuvier 1819), *H. vittatus* (Castelnau 1861), *Hydrocynus brevis* (Gunther 1864), *H. goliath* (Boulenger 1898) and *H. tanzaniae* (Brewster 1986) (see Figure 1.5). The taxonomy of *Hydrocynus* has undergone previous discussion and revision (e.g. Brewster, 1986; Paugy & Guegan, 1989; Skelton, 1990), which can result in misidentification and confusion. Each recognised species, the location of its type locality and taxonomic revisions involving these species will be briefly covered below.

Figure 1.5: Photographs of all the recognised *Hydrocynus* species
From bottom (clockwise): *H. goliath*, *H. vittatus*, *H. forskahlii*, *H. tanzaniae* and *H. brevis*.

H. forskahlii was described by Cuvier (1819) from the Nile River. *H. vittatus* Castelnau (1861), which was described from the Okavango, was synonomised with *H. forskahlii* by Brewster (1986). However, this decision was reversed by Paugy and Guegan (1989), who restored *H. vittatus* to separate species status. In this study, *H. vittatus* will be treated as a separate species following Paugy and Guegan. *H. vittatus* is the best researched species (e.g. Jackson, 1961; Bell-Cross, 1965b; Gaigher, 1970; Kenmuir, 1973). *H. lineatus* (Bleeker 1863) is a synonym of *H. vittatus* (Brewster, 1986). *H. brevis* (Gunther 1864) was also described from the Nile. *H. somonorum* (Daget 1954) is a synonym (Brewster, 1986). *H. goliath* (Boulenger 1898) was described from the Congo system and has *H. vittiger* (Boulenger 1907) as a synonym. *H. tanzaniae* was described by Brewster (1986) from the Ruvu River. This species is restricted to the East Coastal Icthyofaunal Province, where it is confined to the Rufiji-Ruaha drainage system and the neighbouring Ruvu River.

The geographical range of *Hydrocynus* is confined to the warmer tropical rivers of Africa. This widely distributed genus is represented in most of the African continent's freshwater systems of both High and Low Africa and on both sides of the EARS (Figure 1.6). Several of the species occur in sympatry. For example, *H. goliath* and *H. vittatus* occur together in the Congo, while *H. forskahlii*, *H. brevis* and *H. vittatus* inhabit rivers of the Nilo-Sudan ichthyofaunal region. In widely distributed genera, there is often much uncertainty about relationships between species and their taxonomy (Skelton, 1988) as demonstrated above.

Fossils of *Hydrocynus* are comparatively abundant in some sedimentary fossil beds. They consist mainly of shed teeth, which *Hydrocynus* individuals replace throughout their lives (Gagiano *et al*., 1996), and occasionally a portion of jaw. The oldest fossils are known from East Africa (Lothagam, Kenya) from the Late Miocene (Stewart, 2001). However, these fossils cannot be identified to species and so can only be used to set the minimum age of first appearance of the genus.

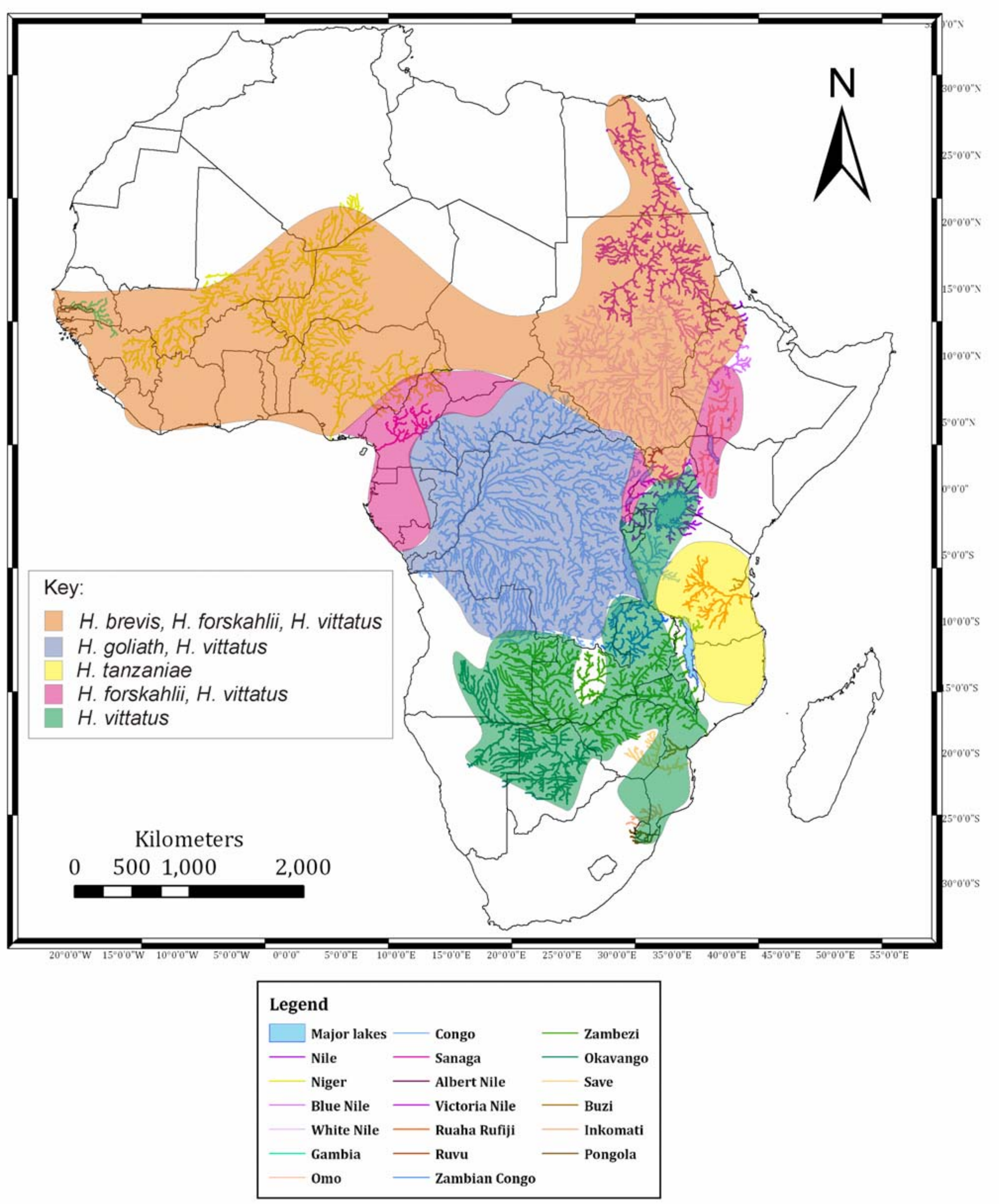


Figure 1.6: Map showing the distribution of *Hydrocynus* species in Africa
Map constructed using ArcMap 9.3 from the HydroSHEDS digital river database (Lehner *et al.*, 2008). Political boundaries are included for geographical context.

1.3.2 *Hydrocynus vittatus*

As *H. vittatus* is the only species that occurs throughout south-central Africa, it was selected as an indicator taxon of drainage evolution in the current genetic study. This species occurs widely across tropical Africa and has an anomalous distribution in south-central Africa (see section 1.3.2; Figure 1.6). *H. vittatus* are charismatic, keystone open-water predators which have a high ecological impact through structuring fish assemblages in the systems where they occur, preying on fish up to 40% of their body length and swallowing them whole, head first (Jackson, 1961; Skelton, 2001). *H. vittatus* is found in the larger rivers and lakes in the area of its distribution and only occurs in swampy habitats if there is an available flow of well-oxygenated water in the form of a large water channel, actively moving away from deoxygenated waters (Jackson, 1961). From studies conducted in the Crocodile River (Kruger Park, South Africa) and in the Upper Zambezi River, it has been established that *H. vittatus* prefers deep in-stream areas and utilises an average depth of approximately 1.32 m (Roux & van Buydner, 2003) and 3.8 m (Thorstad *et al.*, 2002), respectively. *Hydrocynus* prefers warm waters (Skelton, 2001) with the temperature below which it cannot survive being approximately 14.5°C (Roux & van Buydner, 2003). *H. vittatus* is thought to migrate annually up large rivers to spawn (Jackson, 1961). However, this may vary between systems since in a radio-tracking study of *H. vittatus* in the Upper Zambezi (Thorstad *et al.*, 2002) no large-scale seasonal upstream movement was recorded. *H. vittatus* can move long distances (mean distance > 26 km over a six-month period) but individual variation is large (Thorstad *et al.*, 2002). The number of eggs produced for every centimetre of total length of an adult female fish is approximately 7290 (Bowmaker, 1973).

In south-central Africa, *H. vittatus* is distributed in the Zambezi, the Okavango (from where it was initial described from Lake Ngami by Castelnau in 1861), the Zambian Congo Rivers and the rivers in the low-lying coastal systems reaching as far south as the Pongola River (Bell-Cross, 1965a; Gabie, 1965; Balon, 1974; Skelton, 2001) (see Figure 1.7). The Pongola constitutes the southern-most distribution for several tropical species of fish (Gabie, 1965). *H. vittatus* is also found in the Congo, where it is believed that most of the southern freshwater fish have originated (Bell-Cross, 1965a; Gabie, 1965). Bowmaker *et al*. (1978) suggests that this is the drainage system of origin of *H. vittatus* as well, dispersing as opportunity presented itself possibly into the Upper Zambezi from the Kasai. However, no conclusive origin for *H. vittatus* in south-central Africa has been identified through genetic studies. Lake Tanganyika and several of the drainage systems of north and west Africa are also home to *H. vittatus* (Skelton, 2001). However, *H. vittatus* is only locally abundant in Lake Tanganyika (Jackson, 1961). In south-central Africa, *H. vittatus* is notably absent from the Upper Save, Cunene and Upper Kafue Rivers and Lake Malawi (Skelton, 2001) (see Figure 1.8). These distribution anomalies suggest that either local extinctions or recent dispersals have modified the overall distribution of these fish. Overall, they point to recent geomorphological changes of these and neighbouring systems.

The restricted distribution of *H. vittatus* is hypothesised to result from a number of physical barriers which contain the dispersal of this species upstream. Jubb (1952), Bell-Cross (1965a) and Skelton (1994) have identified several possible barriers (numbered as in Figure 1.7): 1) the separation of the Upper and Lower Save system by the Chivirira falls; 2) the separation of the Cunene from the Okavango by a watershed; 3) the physical barrier of the Chasunta Falls between the Kafue and the Middle Zambezi; and 4) the Kapachira Rapids as a possible physical barrier on the Shire River preventing upstream distribution into Lake Malawi. All these barriers are of geological origin. Bell-Cross (1965b) reported that a 4m vertical barrier prevents *H. vittatus* and other species invading the upper reaches of the Kashizi River, a tributary of the Upper Zambezi.

Introductions of fish, by man, into water systems beyond where they would be naturally found could also explain anomalous distributions (Jubb, 1977). Thus, any anomalies in fish distribution should be critically evaluated. A known introduction of *H. vittatus* from one drainage system to another in the study area has occurred from the Kariba Dam to a tributary of the Save River at the Lone Starr Ranch in the Malilangwe Conservation Area, north of Gonare Zhou National Park (A. R. Kaschula, pers. comm.).

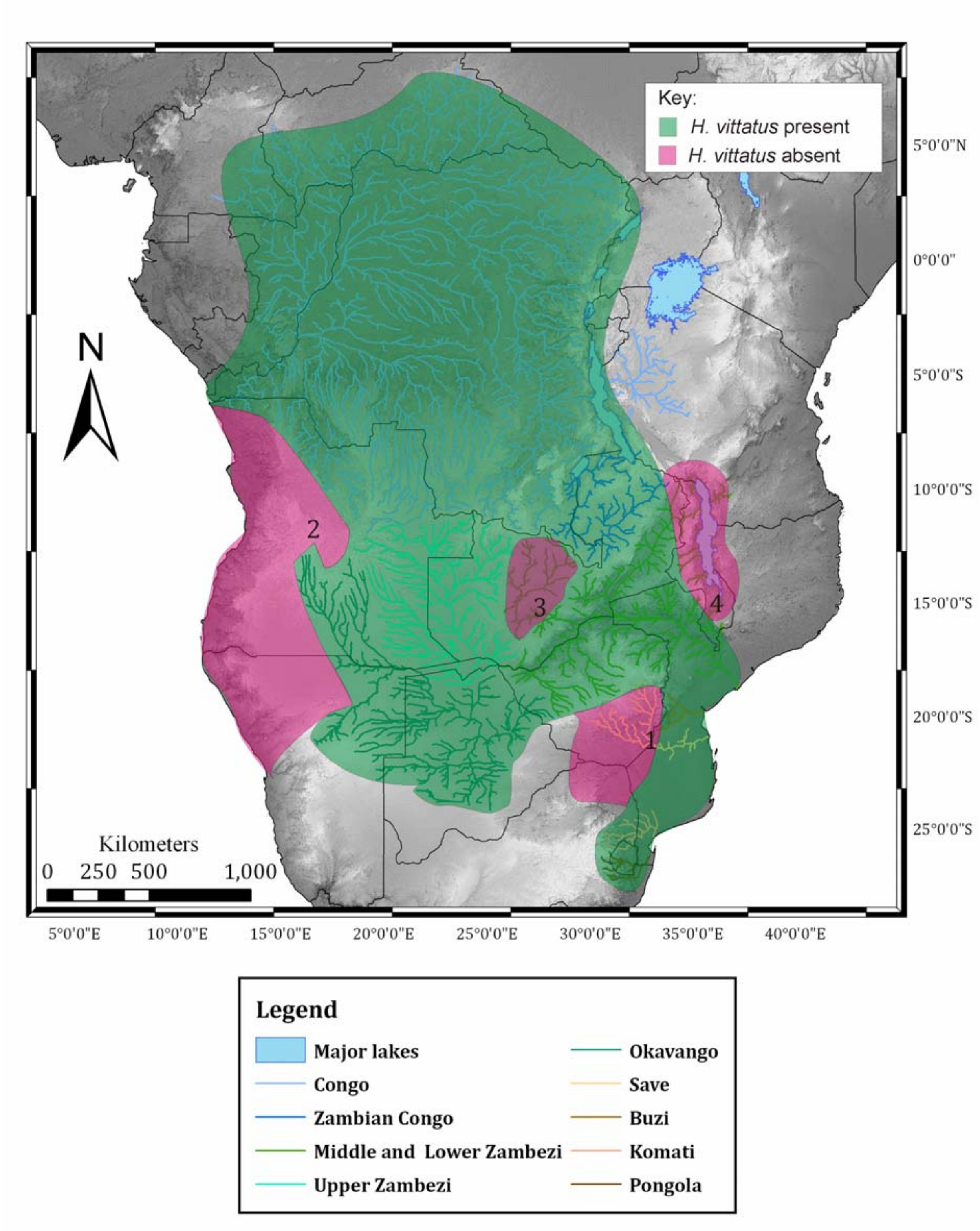


Figure 1.7: Map showing the distribution of *H. vittatus* in south-central Africa
Map constructed using ArcMap 9.3 from the HydroSHEDS digital river database (Lehner *et al.*, 2008) and high-resolution elevation data from the Space Shuttle flight for NASA's SRTM (Jarvis *et al.*, 2006). Political boundaries are included for geographical context. Green = present; Pink = absent. Numbers indicate geological barriers to dispersal discussed in the text.

1.3.3 Hypotheses of Drainage Evolution and the Biogeographical Significance of the Victoria Falls

To account for *H. vittatus*'s anomalous distribution, several hypotheses of drainage evolution have been proposed, in addition to those mentioned in section 1.2. Bell-Cross (1965a) hypothesises that the most likely dispersal route out of the Congo was into Upper Zambezi via the Kasai after it captured the Upper Zambezi's headwaters. Thereafter, *H. vittatus* invaded the Middle Zambezi after the physical barriers formed on the Shire River otherwise it would have reached Lake Malawi. He suggests that this movement into the Middle Zambezi had possibly occurred via the Upper Zambezi. As *H. vittatus* is also absent from the Kafue, Bell-Cross concludes that it must have invaded the Upper Zambezi after the Palaeo-Kafue link with the Upper Zambezi had ceased to exist. Gabie (1965) suggests an alternate route into the coastal systems when the Zambezi, Okavango and associated rivers were joined to the Limpopo.

Several of these hypotheses involve the much debated question as to whether the Victoria Falls (approximately 100m high) is a barrier to downstream fish dispersal or not. Gabie (1965) considers these falls a barrier. Bell-Cross (1965a) suggested a route for *H. vittatus* into the Middle Zambezi via the Victoria Falls yet later stated that a vertical height of 3m is enough to prevent fish movement (Bell-Cross, 1965b). Balon (1974) argues that the Victoria Falls might not be a physical barrier to fish distribution but that the gorge below is ecologically unsuitable for the fish who journeyed over the falls. Jubb (1977) partially disagrees with Balon's hypothesis, stating that fish falling from such a height would be unlikely to survive, and if some did survive, the contrast in environment above and below these falls would make them easy targets for predators. He also suggests that if fish are successfully dispersing from the Upper to the Middle Zambezi, the most plausible route would be via the turbines of the hydroelectric power station that was completed in 1938.

The hypotheses above do not take into account the geomorphological history of the Falls. In order to test the hypothesis that the Victoria Falls is in fact a barrier to distribution, and to test the various hypotheses from the literature of drainage and fish biogeographical histories in south-central Africa outlined above, this study will examine molecular data in both a phylogenetic and a phylogeographic framework.

1.4 Phylogenetics and Molecular Divergence Times

1.4.1 *Hydrocynus* phylogeny

Phylogenies are used to make inferences about evolutionary processes within and among lineages (Avise, 2004). Although several characid molecular phylogenies including species of *Hydrocynus* have been published (van der Bank *et al.*, 2005; Calcagnotto *et al.*, 2005; see Table 1.2), a phylogeny including all recognised species of *Hydrocynus* has not been produced.

Table 1.2: Table of published phylogenies including *Hydrocynus*

Paper	**Species used (no. of individuals)**	**Marker**
van der Bank *et al.* (2005)	*H. vittatus* (7)	Allozymes
Calcagnotto *et al*. (2005)	*H. vittatus* (3) *H. brevis* (1) *H. forskahlii* (1) *H. goliath* (1)	16S, *Cyt b*, fhk, sia, RAG2, TROP

A mtDNA phylogeny will be constructed in this study using *cyt b* sequences from all the recognised species of *Hydrocynus*. MtDNA is a small, circular molecule of DNA generally inherited by offspring through the maternal line without undergoing recombination. Therefore, mtDNA sequence data are used to trace back a species' line of decent through the genealogy representing sequential generations of breeding females (Avise, 2000 & 2004). The *cyt b* region of the mitochondrial genome codes for a polypeptide subunit of the electron transport chain and is part of respiratory complex III which, in teleosts, is generally evolutionarily conserved (Murphy & Collier, 1997). As it is a protein-coding region, c*yt b* evolves at a relatively moderate rate and this allows a species and genus level of resolution, which is ideal for constructing a phylogeny. The knowledge of geological history of the range occupied by a species is a key factor in interpreting phylogenetic patterns in a rational way and these data will be taken into account.

Since a mtDNA phylogeny provides a picture of the maternal lineage through evolutionary time, a nuclear phylogeny is desirable to reconstruct the deeper history of both male and female lines of decent. Nuclear introns have been found to reveal informative genealogical relationships in combination with mtDNA markers (Zhang & Hewitt, 2003; Beheregaray, 2008). Ribosomal protein S7 (*RpS7*) is a nuclear gene with several variable introns (Chow & Hazama, 1998) which can be amplified using primers designed to target the highly conserved flanking exon sequences (Lessa, 1992). *RpS7* intron 2 was chosen for this study as its flanking exons, exons 2 and 3, have been shown to be highly conserved across distant species and universal primers for fish have already been published (Chow & Hazama, 1998) and used successfully in several studies (e.g. Burridge *et al.*, 2006; Morrison *et al.*, 2006; Domingues *et al.*, 2007; Cooke *et al.*, 2009).

A phylogeny shows the relationships between homologous genes as a gene tree with branching events representing events of gene divergence and the resulting relationships are often used to infer relationships between species (Doyle, 1992; Nichols, 2001). This assumes that the gene tree is analogous to the species tree. However, species trees show the real relationships between species with speciation represented by branching events and the species tree can differ substantially from the related gene tree. There are several reasons for this discordance which include introgression, gene duplication and the degree of completeness of lineage sorting (Doyle, 1992; Nichols, 2001). It is possible that introgression, which refers to hybridization between species, could be occurring in *Hydrocynus* given that no research has been conducted on this topic and several species do occur in sympatry in certain rivers. Although this hybridization would not affect non-recombining mtDNA markers directly, one species could carry the mtDNA molecule of another *Hydrocynus* lineage without this being picked up in this study. Incomplete lineage sorting can result in a similar pattern to that created by introgression and is determined by the interaction of mutation rate and effective population size (Avise, 2000). Gene duplication within a genus can result in gene trees constructed from paralogous gene sequences, making them very different to the species tree.

Some molecular genetic studies using mtDNA markers have discovered nuclear copies of mtDNA (numts) (Lopez *et al.*, 1994) in several eukaryotes including humans, cats and rice (e.g. Bensasson *et al.*, 2001; Richly & Leister, 2004). This duplication of mtDNA genes is a concern for any study using mtDNA markers. However, numts appear to be extremely rare in fish (Bensasson *et al.*, 2001; Ivanova *et al.*, 2007) and, if a single product showing the characteristics of the expected authentic product is routinely amplified, the amplification of a numt is considered unlikely (Ivanova *et al.*, 2007).

The discrepancies between gene and species trees highlight the importance of using multiple gene trees, ideally constructed from several genetic markers inherited from both the male and female lines, in reconstructing a picture of the species tree. Despite this, mtDNA trees are thought to result in fairly accurate estimates of divergence times due to this marker's small effective population size (Ne), as a result of being passed on through the maternal line only, and relatively rapid mutation rates across the mtDNA genome (Nichols, 2001; Avise, 2004). The larger the Ne, the larger the error and bias in the timings of the splits between species is thought to be (Nichols, 2001).

1.4.2 Molecular Clocks and Divergence

Inferring the timing of the divergence of lineages from a common ancestor is important to the understanding of evolution within a genus or species and can inform hypotheses of drainage evolution. With molecular data, dates of divergence are calculated using the number of accumulated differences between the DNA sequences of two lineages which is known to be a function of the amount of time since they started on their independent evolution (Bromham & Penny, 2003). This apparent linear relationship between sequence changes and time is termed the molecular clock (Zuckerkandl & Pauling, 1962). These clocks differ between different genes, lineages, species and genera (Posada & Crandall, 2001). Using molecular clocks can help to place evolutionary events from the past in the context of time as well as to shed light on the evolutionary process. The reliability of a molecular clock depends on the accuracy of the genetic distance estimation and on the suitability and precision of the calibration rate. The clock can be calibrated using evidence taken from the fossil record, geological discoveries or age findings from other data using calibrated molecular clocks (Bromham & Penny, 2003). However, the rate of change of homologous sequences can differ vastly between lineages and species, leading to unclock-like behaviour of the changes in the sequences over a molecular phylogeny, and this stochasticity needs to be taken into account when inferring the divergence time (Ayala, 1997; Drummond *et al.*, 2006).

Ideally, a genus- or even species-specific molecular clock using accurately dated geological or fossil calibration points should be used to maximise precision. However, this is generally not possible due to the paucity of appropriate fossils, the inability to identify possible fossils to species level and problems with geological dating. In the case of *Hydrocynus*, no specific substitution rate for *cyt b* currently exists and the fossil record is poor, limited mainly to teeth that cannot be identified beyond the genus level (Stewart, 2001). This significantly limits the use of these fossils to calibrate a molecular clock for the genus. However, based on the available fossils, a minimum Late Miocene age of appearance of the genus can be established (Stewart, 2001) (see section 1.3.1). When no specific molecular clock is available for a particular genus, a general, or universal, rate (e.g. Brown *et al.*, 1979; Berendzen *et al.*, 2008) is often applied to estimate divergence times. Although this method is prone to lower levels of precision, these approximate clocks allow for time estimates that can be used to test biological hypotheses (Bromham & Penny, 2003).

Instead of relying on a strict molecular clock, modelling mutation rate variation over a phylogeny offers a more realistic reconstruction of lineage history. Bayesian methods allow rate changes between lineages in a tree based on a specified model (Bromham & Penny, 2003). BEAST (Bayesian Evolutionary Analysis by Sampling Trees) 1.4.8 package (Drummond & Rambaut, 2007) allows a relaxed clock method to be used within a Bayesian Markov chain Monte Carlo

(MCMC) framework. This "relaxed phylogenetics" method incorporates uncertainty into the estimation of the tree, accommodating rate heterogeneity that has accumulated across gene trees, and so does not require a strict molecular clock (Drummond *et al*., 2006). It allows for the co-estimation of the phylogeny and the divergence times. In this study, the times of divergence among lineages are estimated using BEAST (Drummond & Rambaut, 2007).

1.5 Phylogeography

Phylogeography can be defined as the geographical distribution of genealogical lineages, and this knowledge of distribution patterns, along with that of mutation rates, can be used to provide insight into the history and biogeographical range of populations (Avise, 1987 & 2004). Phylogeographic studies provide a convenient way to investigate levels of diversity within and among populations, as well as understand the historical context of a population's structure.

The phylogeographic approach is particularly informative with regard to species with close phylogenetic relationships (Avise, 2000 & 2004). It makes use of data on the distribution of evolutionary relationships in space and time by reconstructing gene trees, analyzing the demographic history of a species and taking into account the environmental features that influence the distribution patterns of species. Thus information from a broad variety of fields such as historical geography, geology and molecular biology is used, making this a highly informative, interdisciplinary field (Avise, 2000; Russell *et al.*, 2007). Appropriate sampling of both genes and individuals can allow for the testing of biogeographic hypotheses and the inference of processes underlying the origin and distribution of a genus or species (Beheregaray, 2008).

Ideally phylogeographic analyses should be based on nuclear data in addition to mtDNA data but mtDNA phylogeographic studies alone give an insight into female population history (Avise, 1995 & 2009). The majority of studies in the field of phylogeography to date make use of mtDNA sequence analyses (e.g. Hrbek *et al.*, 2004; Katongo *et al.*, 2005; Modolo *et al.*, 2005;

Burridge *et al.*, 2007). MtDNA has several properties which make it an ideal marker for these studies (Avise *et al.*, 1987) (see section 1.4). This small, circular molecule of DNA is generally maternally inherited without undergoing recombination. Therefore, sequence data from this molecule can be used to trace back a species' maternal line of descent (Avise, 2000 & 2004). Its smaller effective population size makes it more susceptible to the effects of bottlenecks and genetic drift. It has a high copy number in cells, which makes it easy to amplify (Avise, 2000). The mitochondrial genome accumulates mutations at a relatively rapid rate compared to single copy nuclear DNA, which allows for variation between populations over a relatively short period of evolutionary time (Avise, 2000). For example, Brown *et al.* (1979) showed that the rate of sequence evolution in higher primate mtDNA was approximately 2% per million years. This fast rate is thought to be due to oxidative damage and intracellular repair enzyme deficiencies (Avise, 1987). All the above properties allow for relationships to be inferred between populations in a phylogeographic context. The non-coding CR in mtDNA initiates the processes of replication and transcription (Avise, 2004). The faster the mutation rate of DNA sequences, the easier it is to resolve closely related clades (Modolo *et al.*, 2005). As the CR is the fastest-evolving region within the mitochondrial genome, it is used widely to determine genetic diversity levels and reconstruct population histories within species (Lee *et al.*, 1995) and in phylogeographic studies (Avise, 2000). The CR has three main regions: two hyper-variable regions at the 5' and 3' ends flanking a relatively conserved middle region.

Many phylogeographic studies have been conducted on both marine and freshwater fish (e.g. Bowen & Grant, 1997; Kreiser *et al.*, 2001; Katongo *et al.*, 2005) and have yielded interesting results. A study on regional sardine (*Sardinops* spp.) populations in the Pacific Ocean reveals a high level of geographical structuring and indicates that populations have undergone an unexpectedly recent (Pleistocene) colonisation around continental margins (Bowen and Grant, 1997). Kreiser *et al.* (2001) show that populations of freshwater killifish (*Funudulus zebrinus*) in North America were geographically structured into northern, central and southern populations with a large divergence between the northern/central and the southern populations. In a study involving the haplochromine cichlid genus *Pseudocrenilabrus* in south-central Africa, Katongo *et al.* (2005) show geographic structuring and speciation resulting in two novel mtDNA lineages, which were identified in Lake Mweru and the Lunzua River, a tributary of Lake Tanganyika.

The historical changes of landscapes and their rivers are intertwined with the history of the freshwater fish that inhabit them, often resulting in phylogeographic structuring of extant populations in relation to landforms (Bermingham & Avise, 1986; Waters *et al.*, 2007). The phylogeographic patterns of a species can be affected by a number of factors such as the ability to disperse, barriers to dispersal and the life history of the species under consideration (Burridge *et al.*, 2008). Judging from its wide distribution and its ability to travel long distances, *H. vittatus* appears to be highly dispersive. However, this ability is restricted to fast-flowing, well-oxygenated water (Jackson, 1961) as discussed in section 1.4, which has important implications for how *H. vittatus* is utilised as biotic indicator of landscape evolution. In this study, the CR genetic data from the Vittatus complex is analysed in a phylogeographic context in order to explore drainage changes in south-central Africa.

1.6 Research Objectives

In the context of the freshwater environment, the evolutionary genetic history of fish populations can be used to understand relationships, both past and present, between different bodies of water. In this study, both the biogeography and phylogeography of the genus *Hydrocynus* is hypothesised to reflect past changes in drainage topology and its causes in south-central Africa over evolutionary scales of time and space. Thus, the objectives of this study are to:

1) construct a phylogeny of the *Hydrocynus* genus using *cyt b* to elucidate the relationships among the species and attempt to date the divergences between these lineages;
2) quantify the level of mtDNA control region variation in *H. vittatus* and its

closely related lineages from various parts of its current biogeographical range;

3) analyse these data in a phylogeographic framework to describe spatial and temporal patterns of genetic variation in this species among the different geographic populations sampled; and

4) use these results to infer past demographic events within *H. vittatus* and test hypotheses of drainage evolution in south-central Africa that have been previously proposed.

Chapter 2

Hydrocynus Phylogeny

2.1 Introduction

In this chapter, a mtDNA phylogeny will be constructed using *cyt b* sequences representative of all the recognised species of *Hydrocynus*. The timing of the divergences between these lineages will be estimated to reconstruct the spatial history of *Hydrocynus* in the context of time.

2.2 Materials and Methods

2.2.1 Sample Collection

Fin clips or muscle tissue were collected from 240 *Hydrocynus* individuals (both by the author and those acknowledged) sampled from 12 drainage systems in the study area (Figure 2.1) (see Appendix 1 for a list of samples, localities and their respective haplotypes). Samples were preserved in 96% (v/v) ethanol and stored at room temperature or 4°C until DNA extraction. A subset of 90 of these samples was selected for PCR amplification and sequencing to create a phylogeny for *Hydrocynus*. Selection was based on morphological identification to species of the specimen from which the sample was taken, in order to represent all recognised species in the phylogeny.

2.2.2 DNA Isolation, Amplification and Sequencing

DNA extraction was performed using two protocols depending on the quantity of tissue obtained. When a large piece of tissue was available, extraction was performed on a small sub-piece of tissue (approximately 2mm^3) using the chelex extraction protocol described by Walsh *et al.* (1991). Chelex resin facilitates the binding of possible PCR inhibitors. Each sub-sampled piece of tissue for extraction was washed in 500µL 1x salt-tris-EDTA buffer and then placed in a 1.5ml eppendorf tube containing 500µL of 5% chelex-100 resin mix (Bio-rad). Proteinase K (Fermentas) was added (13.5 U) and the sample was incubated at

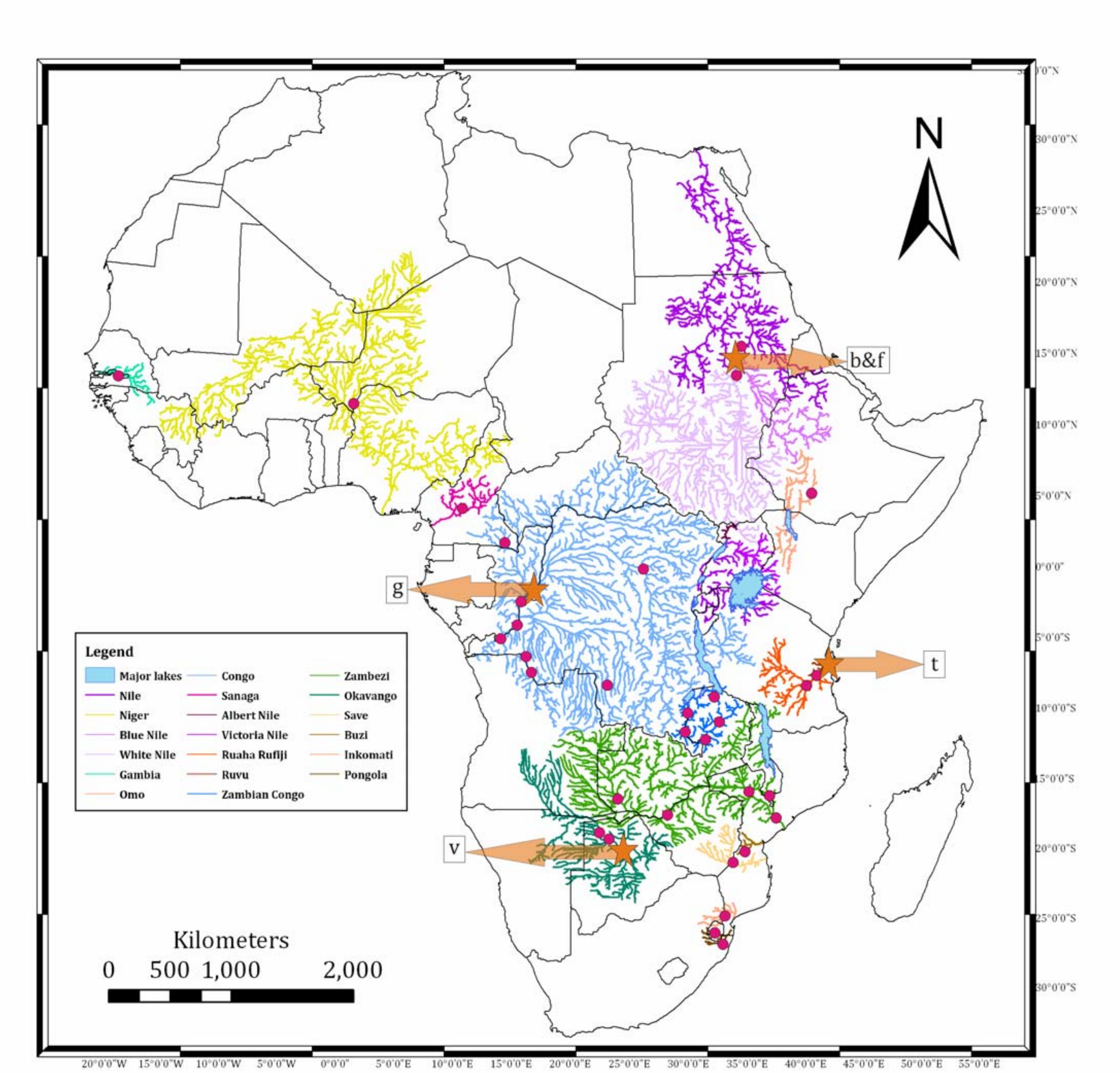


Figure 2.1: Map of Africa indicating rivers sampled for species of *Hydrocynus*

Map constructed using ArcMap 9.3 from the HydroSHEDS digital river database (Lehner *et al.*, 2008) and high-resolution elevation data from the Space Shuttle flight for NASA's SRTM (Jarvis *et al*., 2006). Sampling points indicated by pink circles. Type localities are indicated by an orange star and labelled with the first letter of the species name: v = *H. vittatus*, g = *H. goliath*, t = *H. tanzaniae*, f = *H. forskahlii* and b = *H. brevis*.

55°C overnight, with occasional vortexing. When solid particles of tissue remained, the sample was treated with collagenase enzyme (50mg/µL) until all tissue had been completely digested. To deactivate the digestive enzymes, the sample was then incubated for 10 min at 95°C, vortexed briefly and centrifuged for 1 min at 13 000rpm to pellet the chelex resin. The supernatant was transferred to a new tube containing 250µL of the 5% chelex resin mix. The resulting supernatant was then processed through a Wizard® SV Gel and PCR Clean-Up System column (Promega) as per the manufacturer's instructions, the final DNA solution being eluted in 100µL of nuclease-free water. This post-extraction clean-up was carried out in order to remove any remaining PCR inhibitors within the extracted DNA. Where only a small piece of tissue was available, extraction was performed using the QIAamp DNA Micro Kit (Qiagen) as per the manufacturer's instructions in order to optimise DNA yields.

A combination of primers designed in this study and modified universal PCR primers were used to amplify the *cyt b* region (Table 2.1). Primers designed in this study were produced from an alignment of available *Hydrocynus* sequences and sequences of closely related characiforms from Genbank (Benson *et al.*, 2005) and relied principally on partial *cyt b* sequences from Calcagnotto *et al.* (2005). To obtain the final *cyt b* fragment, the amplification proceeded in two parts. Primer pairs L14724HycF2 or L14990FishF and H15494HycR2 were used to target the first half and L15408Hyc and CRFish-RHyc, HycGR2, HycR3 or H15919HycR, were used to amplify the second as partially overlapping amplicons (Figure 2.2). The large number of reverse primers used was a result of the variability between *Hydrocynus* species at the region initially selected for primer binding, based on the partial Genbank sequences from Calcagnotto *et al.* (2005) and related characiform fish.

Table 2.1: *Cyt b* primers and their sequences (5' to 3')

Primer	**Sequence (5' to 3')**	**Source**
H14724Hyc2	GACCTATGACTTGAAAAACCATCGTTG	Modified from Irwin *et al.* (1991)
L14990FishF	ATYTCAGCATGATGAAAYTTTGGTM	Modified from Irwin *et al.* (1991)
H15494HycR2	AARTATCATTCTGGCTTRATATGYGG	Modified from Irwin *et al.* (1991)
L15408HycF	GAAACAGGCTCCAACAACC	Modified from Irwin *et al.* (1991)
H15915HycR	TTAACCTTCRATCTTCGGATTACAAGAC	Modified from Irwin *et al.* (1991)
HycGR2	GTATCTCTTGATGTTATGTCC	Designed for this study
HycR3	AGAATGCTTTATGTCTACC	Designed for this study

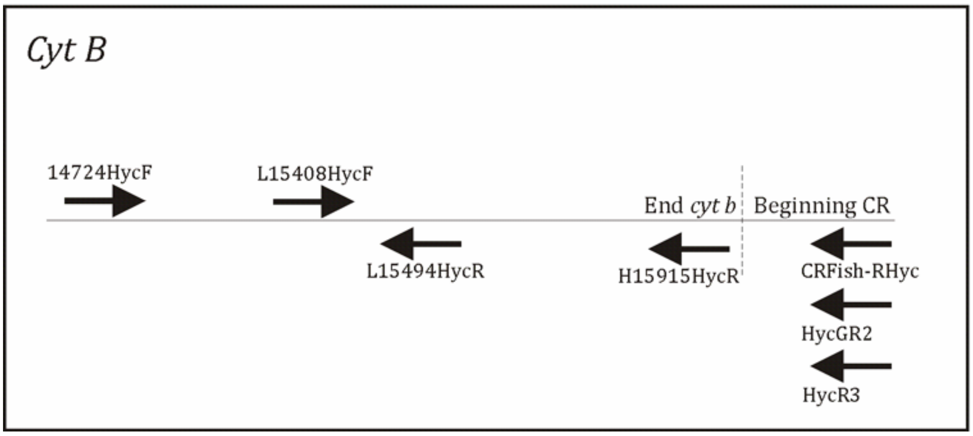


Figure 2.2: Schematic representation of the positions of primers for amplification of *cyt b* of *Hydrocynus*

For all primer pairs used to amplify the *cyt b* region, the PCR reaction mixture per sample (total volume 50µL) was composed of 0.5U SuperTherm Taq (Southern Cross Biotech), 200µM dNTP's, 1x buffer, 2mM $MgCl_2$, 0.5µM of each primer, 1mg/µL BSA and 10-50ng template DNA. Approximately 20µL of mineral oil was added to each sample as an overlay. Reactions were carried out on a Programmable Thermal Controller-100 (MJ Research, Inc.) with the following PCR conditions: 94°C for 5 min; 5 cycles of amplification composed of 94°C for 40 sec, 48°C for 50 sec and 72°C for 1 min; 30 cycles of amplification composed of 94°C for 40 sec, 50°C for 50 sec and 72°C for 1 min; and a final extension step of 72°C for 10 min.

PCR products were electrophoresed on 2% (v/v) agarose gels (1x TBE buffer) with O'Gene Ruler 1kb DNA Ladder (Fermentas). The target band was excised from the gel and purified using a Wizard® SV Gel and PCR Clean-Up System column (Promega) as per the manufacturer's instructions. The purified DNA fragments (10 – 20ng of purified PCR product) were then sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). Reactions were carried out on an ABI 2700 thermocycler (Applied Biosystems) with the following conditions: 25 cycles of 94°C for 30 sec, 50°C for 5 sec and 60°C for 4 min. To aid the removal of unincorporated nucleotides from the final sequence product, 0.5% SDS was added to the sequencing reactions followed by 10 minutes at 95°C and 5 min at 25°C. Sequencing reactions were run on an ABI 3100 sequencer (Applied Biosystems).

In order to create the nuclear data set for analysis, amplification of the S7 intron 2 nuclear region was attempted. Universal PCR primers for the *RpS7* intron 2 region were used (Chow & Hazama, 1998) (Table 2.2).

Table 2.2: *RpS7* intron 2 primers and their sequences (5' to 3')

Primer	**Sequence (5' to 3')**	**Source**
S7RPEX2F	AGCGCCAAAATAGTGAAGCC	Chow & Hazama (1998)
S7RPEX3R	GCCTTCAGGTCAGAGTTCAT	Chow & Hazama (1998)

The PCR products were electrophoresed, excised, purified and sequenced as for the *cyt b* fragment. However, this intron did not amplify any usable sequence with the published primers. Cloning of a selection of amplified products using the pGEM®-T Easy Vector System (Promega) as per the manufacturer's instructions was attempted to overcome this problem. The ligated vector DNA was transformed into DH5α competent cells. However, the resulting sequences had high background. Due to time and budget constraints, the S7 marker could not be further investigated in this study and will, therefore, not be discussed any further.

2.2.3 Data Analyses

Sequence Editing and Alignment

DNA sequences were edited using Chromas 2.13 (Technelysium, 2006). The ClustalW feature (Thompson *et al.*, 1994) in Bioedit 7.0.5.2 (Hall, 1999) was used to create alignments. An alignment of all 90 *Hydrocynus cyt b* sequences was created for phylogenetic analyses. The sequences were checked by toggle translation into amino acids to ensure that they represented functional protein. The fact all the sequences represent functional proteins and contain no indels makes it unlikely that any of the sequences represent numts (Sorenson & Quinn, 1998; Ivanova *et al.*, 2007) (see section 1.4). Alignments were adjusted by eye where necessary (see Appendix 2 for the 823 bp *cyt b* haplotype alignment). The identity of all sequences was checked using the BLASTN search in the NCBI nucleotide database on Genbank (Benson *et al.*, 2005).

Previous *cyt b* sequence data was available for only *H. vittatus* (Genbank accession numbers: AY791404; AY791410), *H. brevis* (AY791400) and *H. forskahlii* (AY791403) (Calcagnotto *et al.*, 2005). However, the locality of collection was unknown for AY791410 (*H. vittatus*) and searching the collection database of the institution at which the voucher specimen is deposited revealed no information on its provenance. This sequence was therefore excluded. The *H. forskahlii* sequence grouped with *H. vittatus* in the analyses. As no voucher specimen appears to exist to check its identity, its scientific value is questionable as it is impossible to find out any information about the sample (especially its geographical provenance and circumstances of collection). It was excluded as a case of possible misidentification.

Phylogenetic Analysis

Paup 4.0b10 (Swofford, 1998) and ModelTest 3.06 (Posada & Crandall, 1998) were used to establish the best-fit model of nucleotide substitution for the sequence data sets as determined by the hierarchical likelihood ratio tests (hLRT). The establishment of the best-fit model for the data is an important step since accurate phylogenetic inferences for a particular data set are dependent on the use of an appropriate DNA evolution model (Huelsenbeck *et al.*, 2001). The

General Time Reversible (GTR) model (Tavaré, 1986) plus Gamma (G) was chosen for the *cyt b* data. Lists of unique sequences (haplotypes) for the *cyt b* data set were generated using Collapse 1.2 (Posada, 1999) (see Appendix 1).

Phylogenies of *Hydrocynus* using the *cyt b* data set were constructed in BEAST 1.4.8 package (Drummond & Rambaut, 2007), MrBayes 3.1.1 (Huelsenbeck & Ronquist, 2001) and PAUP 4.0b10 (Swofford, 1998), using the GTR model parameters for each data set. Both BEAST and MrBayes use Bayesian inference and a MCMC sampling procedure to reconstruct a phylogeny by estimating the probability distribution given sequence data (Huelsenbeck & Ronquist, 2001; Drummond & Rambaut, 2007). These algorithms weight trees in proportion to their posterior probability, such that a branch with a posterior probability close to 1 is considered well supported, whereas a value close to 0 is considered very weakly supported.

In BEAST, all starting trees were constructed using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) (Sokal & Sneath, 1963). The GTR model of nucleotide evolution and its parameters were specified and the uncorrelated lognormal relaxed molecular clock was used to allow for independent substitution rates on different branches. Short runs of 500 000 generations were performed to optimise the operators, which allows the algorithm to sample the target distribution faster and more efficiently, using the suggestions of the previous run in the following run until no more suggestions were made. A long run of 50 000 000 generations was then performed, logging parameters every 1 000 generations. Burnin was set to the first 5 000 logged trees. All other priors and operators were kept at the default settings. The methods of dating estimation using BEAST are presented at the end of this section.

In MrBayes, starting trees were randomly selected. Four simultaneous Markov chains were run for 5 000 000 generations with trees sampled after every 50 generations, giving a total of 100 001 trees. Burnin was set to the first 25 000 trees (10%).

The Maximum Parsimony algorithm infers phylogenies from discrete-character data by searching for the trees of minimum length, which require the least amount of evolutionary change to account for the data. As implemented in PAUP, the program searches for the most parsimonious trees. A heuristic search was conducted in order to estimate the most likely topology. The starting tree was constructed using stepwise addition and the tree-bisection–reconnection (TBR) method was used as the branch swapping algorithm. The number of bootstrap replications conducted to estimate the level of support for the resulting topologies was 10 000.

As different tree building algorithms can infer different tree topologies and thus result in different interpretations of the data, using a combination of programs with different statistical algorithms should give a better representation of the true phylogeny. *Microlestes acutidens* (Genbank accession number: AY791418.1) and *Ladigesia roloffi* (AY791417.1) were selected as outgroups due to their close genetic relationship to *Hydrocynus* identified by Calcagnotto *et al.* (2005). Taxa were accepted as belonging to a certain species based on their relationship to known sequence data from that species from known localities, where available, and to sequences from topotypical material (obtained from the type locality). For example, *H. vittatus* from the Okavango Delta is considered topotypical (Brewster, 1986; Paugy & Guegan, 1989; see Figure 2.1).

In order to estimate the ages of the lineage splits of *Hydrocynus*, a relaxed phylogenetics method was implemented in BEAST 1.4.8 (Drummond & Rambaut, 2007). This method employs a relaxed molecular clock that co-estimates the tree and dates of divergence of the lineages, given the stipulated model of nucleotide substation (GTR). An uncorrelated lognormal relaxed molecular clock prior with a Species Birth-death tree prior was used to estimate the timing of divergences between lineages (Drummond *et al.*, 2006). Short runs of 100 000 generations were performed to optimise the operators, which allows the algorithm to sample the target distribution faster and more efficiently, using the suggestions of the previous run in the following run until no more suggestions were made. A long run of 50 000 000 generations was then performed, logging parameters every 1 000 generations. All other priors and operators were kept at the default settings. Burnin was set to the first 5 000 logged trees. The minimum age of appearance of *Hydrocynus* (approximately 12 Ma – Stewart, 2001) was used as a calibration point of the root of the tree with a lognormal distribution. As there is no *Hydrocynus* specific *cyt b* evolutionary rate, a generalised teleost *cyt b* substitution rate of 0.76 - 2.2% per million years was used, as applied by Berendzen *et al*. (2008) as a uniform prior. All other priors and operators were kept at the default settings.

Nucleotide and Haplotype Diversity and Genetic Divergence between Lineages

Genetic diversity was characterized by the haplotype and nucleotide diversities estimated in Arlequin 3.11 (Excoffier *et al.*, 2005) and genetic distances within and between lineages estimated in Mega 4.0 (Tamura *et al.*, 2007) using the Kimura 2-parameter model (Kimura, 1980). Rate heterogeneity among sites was corrected for with the gamma shape parameter.

Population Structure and Demographic History

A measure of genealogical sorting is important for assessing the amount of differentiation between and within species. The genealogical sorting index (gsi) (Cummings *et al.*, 2008) quantifies the degree of common ancestry of groups of individuals on a rooted phylogenetic tree. This statistic was used to quantify the common ancestry of *cyt b* sequences in an interspecific context. A gsi value of 1 represents a monophyletic group, while a group of completely mixed genealogical ancestry will have a gsi value of 0. This is an informative statistic in assessing the role of river geomorphology in determining the levels of differentiation among clades since assignment to groups can be based on sampling locality, giving insight into the history of different populations. This analysis was run on the gsi website (Bazinet *et al.*, in prep.). The groups in the assignment file for the *cyt b* gsi analysis were set according to species and evolutionary lineage. A Newick version of the BEAST tree was used in the analysis. The number of bootstrap replications conducted to estimate the gsi for the assigned groups was 10 000.

To test whether extant drainage patterns have influenced the population structure, an analysis of molecular variance (AMOVA) was calculated in Arlequin 3.11 (Excoffier *et al.*, 1992). This analysis calculates the fixation index (ΦST) (Wright, 1951) which uses haplotype diversities to estimate levels of genetic diversity and differentiation between contrasted populations. ΦST analysis partitions the total genetic variance into among- and within-population components, with population specific ΦST's calculated.

Tajima's *D* (Tajima, 1989) and Fu's *Fs* (Fu, 1996) statistics test for deviations from the neutral theory model for a population of constant size (Kimura, 1968). The neutral theory model assumes that most mutations are neutral and that sampling of genes occurs from a freely mixing population which is at equilibrium. Tajima's *D* examines whether the number of polymorphic sites correlates with the average number of nucleotide differences in the data set. Fu's *Fs* considers, given the number of observed haplotypes, whether it is probable that a larger number exists (Fu, 1997). In the event of population growth or background selection, larger numbers of low-frequency alleles would be expected to be encountered if the population size was constant, and this would result in a large negative value for the neutrality test. Fu's *Fs* is particularly sensitive to the signals of recent population expansions (Fu, 1997).

Mismatch distributions (Li, 1977; Harpending, 1994; Rogers, 1995), the distributions of the observed number of differences between pairs of haplotypes, were used to quantify historical changes in population size in the *Hydrocynus* lineages. A unimodal ('smooth') distribution indicates recent population growth compared to a multimodal ('ragged') distribution which indicates demographic stability. Harpending's 'raggedness' index (*r*) (Harpending, 1994) was used to quantify the "smoothness" of the distribution. Mismatch distributions were only calculated for lineages that showed a significantly negative Tajima's *D* (Tajima, 1989) and/or Fu's *Fs* (Fu, 1996). Tajima's *D*, Fu's *Fs* and the mismatch analyses were only performed on sample groups with four or more individuals. The AMOVA, Tajima's *D*, Fu's *Fs* and the mismatch analyses were carried out in

Arlequin 3.11 (Excoffier *et al.*, 2005) with 10 000 bootstrap replications conducted for each analysis.

2.3 Results and Discussion

The results from the mtDNA *cyt b* sequence data analysed in a phylogenetic context will be presented and discussed below.

2.3.1 Data and Haplotypes Summary

The 823 bp *cyt b* fragment obtained from 90 individuals yielded 42 haplotypes (listed in Appendix 1). The General Time Reversible (GTR) model (Tavaré, 1986) plus Gamma (G) was chosen for the *cyt b* data (Appendix 3.1). This model accounts for variable base frequencies and allows substitution rates to differ among sites. The *cyt b* region was found to be C and A-T rich but G poor, which is the same general pattern found in a study on *cyt b* evolution involving 81 genera of fish (Johns & Avise, 1998), and a higher proportion of transitions were also found to occur. The Gamma shape parameter (α) specifies the range of rate variation between pairs of nucleotides. When α is small (< 1) it indicates that most of the sites evolve at a slow rate but a few sites evolve at a moderate or fast rate. As α increases (> 1), the majority of the sites evolve at moderate rates while a few evolve at fast and slow rates. When α equals infinity then all the sites evolve at an equal rate (Posada & Crandall, 2001). The Gamma shape parameter of 0.1691 indicates extreme rate variation across the data.

Haplotypes shared between individuals from different drainage systems and different rivers within the same system are discussed below (Appendix 1; Figure 2.3-2.5). Haplotype 40 was shared between *H. brevis* from the Gambia, Niger and the Nile Rivers. This indicates a relatively recent connection and/or gene flow across north-west Africa. Roberts (1975) hypothesises that the Gambia and Senegal Rivers nearly dried out during the last dry climatic period (approximately 27 - 12 ka) and that the current Gambian fish fauna is a result of colonisation from the Senegal River, which was re-colonised by fish from the Niger River during the last wet climatic period (approximately 12 – 8 ka) (Lévêque, 1997). A recent link between the Niger and the Nile is also possible

(Reid, 1996). This is consistent with the haplotype shared between the three systems. Several species are shared between the Nile and drainage systems across west Africa (Roberts, 1975; Reid, 1996), which points to relatively recent faunal exchanges. Broader sampling of these icthyofaunal regions and further analysis would help to provide more detailed resolution on the relationships between *Hydrocynus* species across north Africa.

Haplotype 4 was shared by *H. vittatus* individuals from the Lower Congo and Kwango Rivers while haplotype 24 was shared by Group A individuals from the Congo River (near Kisingani) and the Lulua River. It is significant that the collecting localities for these individuals span over 4 000 kilometres. This indicates that populations of *Hydrocynus* species in the Congo system have been connected in at least the relatively recent past across most of the drainage system. This is somewhat surprising due to the large size of the basin and the known barriers that would appear to restrict dispersal of fish (e.g. Portes d'Enfer) (see section 1.2.1 and Chapter 4).

Haplotype 21 was shared between individuals from Lakes Mweru and Bangweulu as well as the Luapula River and haplotype 26 was shared by individuals from Lakes Mweru and Bangweulu only. These shared haplotypes across the Zambian Congo support Bell-Cross's (1965a) hypothesis that the waterfalls on the Luapula River do not restrict fish movement, at least in the recent evolutionary past (see section 1.2.1 and Chapter 4).

Haplotype 6 consisted of Okavango and Upper Zambezi individuals. This confirms the intermittent connection between the Okavango and Upper Zambezi systems (Bell-Cross, 1965a & 1972; Roberts, 1975).

Haplotype 8 was composed of *H. vittatus* individuals from coastal populations south of the Lower Zambezi and an individual from the Middle Zambezi. The presence of the same haplotype in this large area and the fact that this haplotype is the most abundant in the coastal populations is consistent with recent colonisation of the coastal systems from the Middle and Lower Zambezi region.

2.3.2 Phylogenetic Analysis

All three programs produced trees with consistent major branching orders that exhibit a strong topological congruence. This lends support to the phylogenetic relationships reconstructed within *Hydrocynus*. The majority of nodes are recovered with high bootstrap and posterior probability supports (Figure 2.3-2.5). For the 50% majority-rule consensus tree constructed in PAUP, 546 characters of the 823 character alignment were found to be constant and 39 of the variable characters were parsimony uninformative, leaving a total of 238 parsimony informative characters. The Consistency Index was 0.598 and the Retention Index was 0.861.

Interestingly, ten unique evolutionary lineages were identified in all three trees, with high support (posterior and bootstrap) even though only five species are recognised. Lineages not associated with previously described species are designated by letters to distinguish them (Figures 2.3-2.5; Table 2.3). Group A is composed of samples from the Kwango, the main Congo channel upstream of Kisingani and the Lulua River (a tributary of the Kasai River). Group B consists of two samples from the Lufubu River (a tributary of Lake Tanganyika) collected alongside a series of *H. vittatus*. Samples from the Dja River, Lake Mweru and Lake Bangweulu formed Group C. Group D is composed of samples from the Luapula River, Lake Mweru, Lake Bangweulu and the Chambeshi River. Groups B, C, D and *H. vittatus sensu stricto* (s.s.) are designated the Vittatus complex in this study and this complex is analysed in a phylogeographic context and their relationships discussed in Chapter 3. All the *Hydrocynus* lineages and their locations are summarised in Table 2.3.

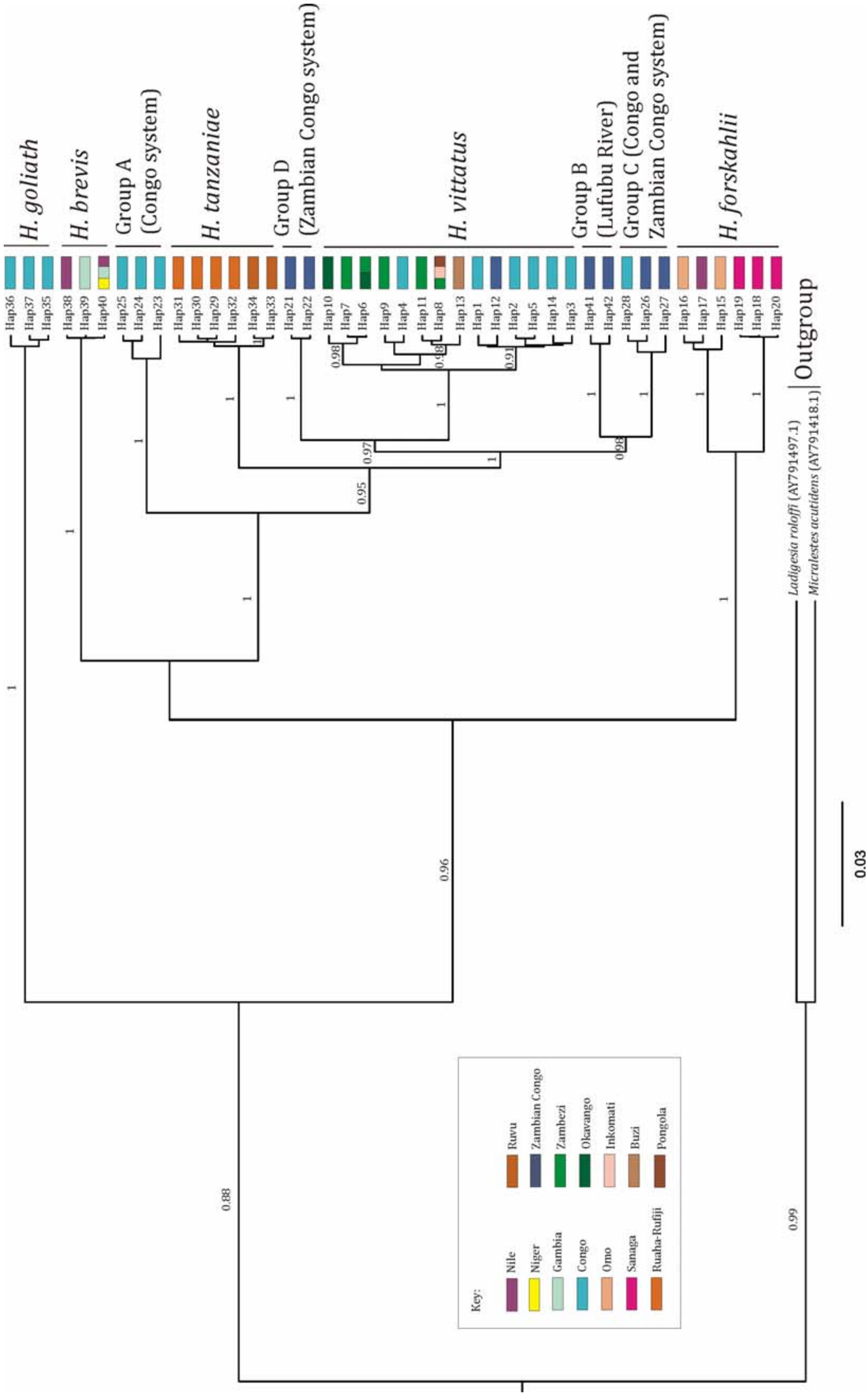


Figure 2.3: Bayesian tree of the *cyt b* sequence data of *Hydrocynus* produced in BEAST, using the GTR parameters specified by Modeltest

Only posterior probabilities over 0.70 are shown. Letters denoting lineages not associated with previously described species as given in the text. The colour key indicates the drainage system(s) represented in each haplotype.

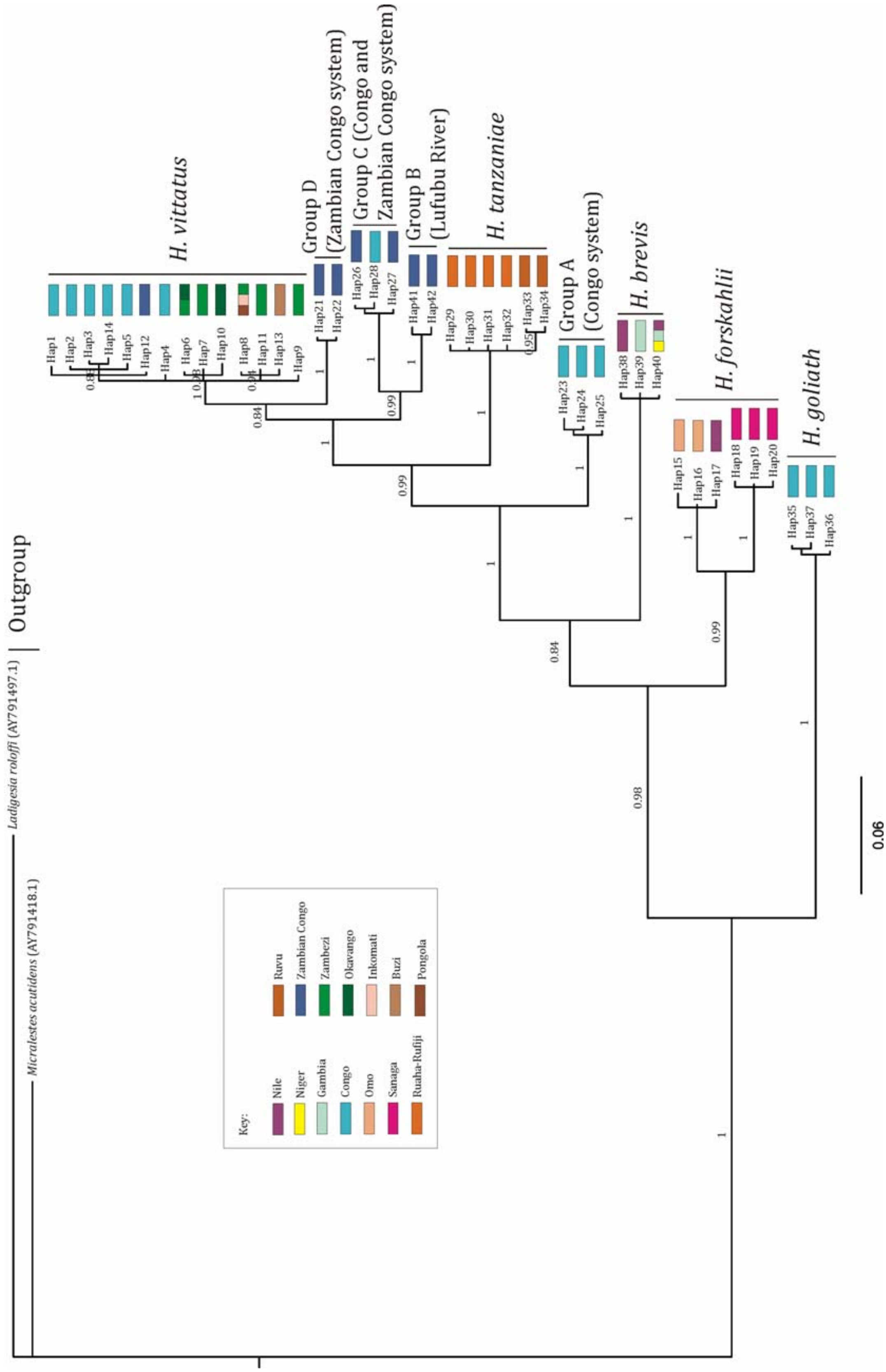


Figure 2.4: Bayesian tree of the *cyt b* sequence data of *Hydrocynus* produced in MrBayes, using the GTR parameters specified by Modeltest

Only posterior probabilities over 0.70 are shown. Letters denoting lineages not associated with previously described species as given in the text. The colour key indicates the drainage system(s) represented in each haplotype.

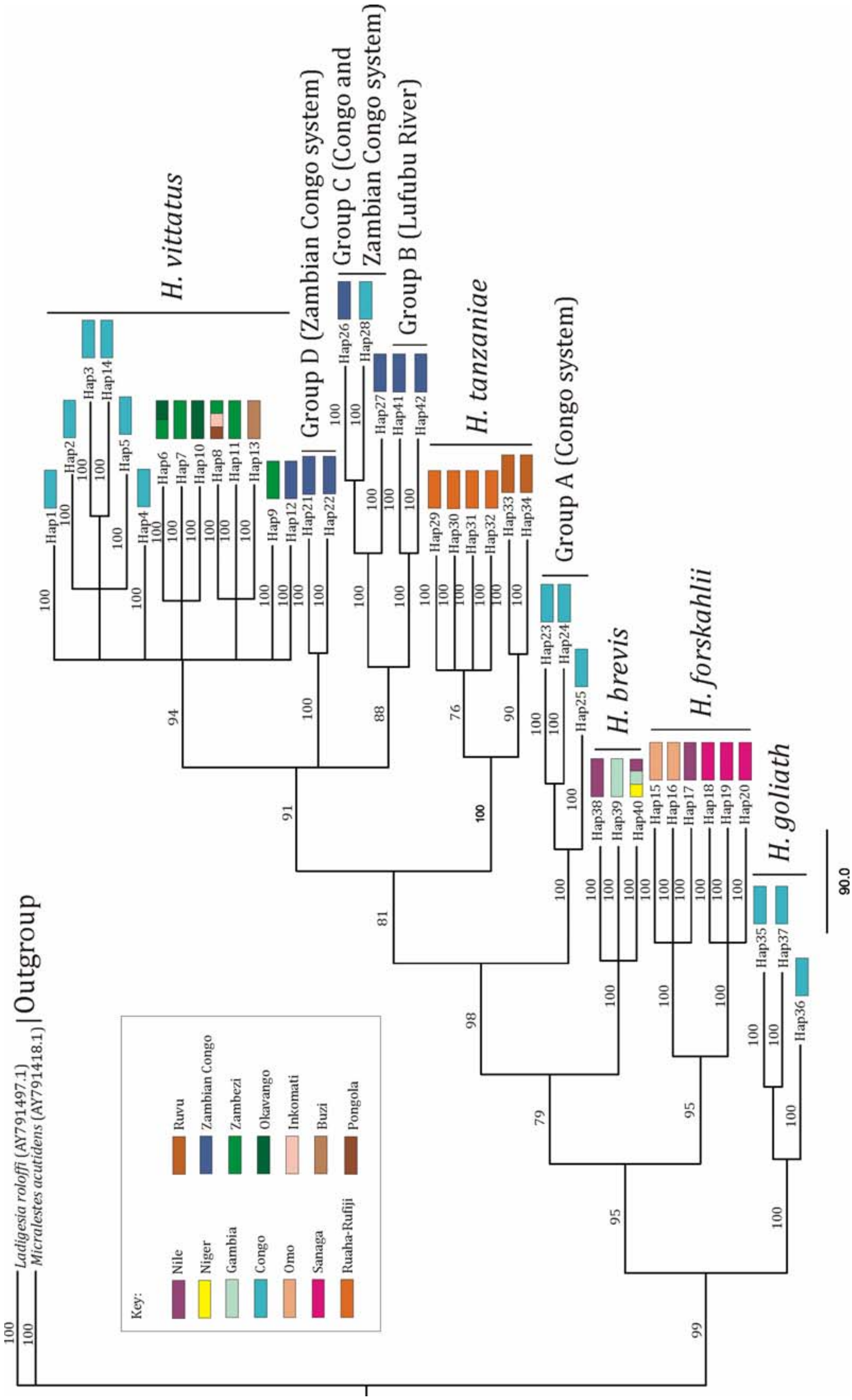


Figure 2.5: Maximum parsimony tree of the *cyt b* sequence data of *Hydrocynus* produced in PAUP, using the GTR parameters specified by Modeltest

Only bootstrap values over 50 are shown. Letters denoting lineages not associated with previously described species as given in the text. The colour key indicates the drainage system(s) represented in each haplotype.

Table 2.3: Summary table of the lineages recovered in the phylogenetic trees of the *cyt b* data

Lineage	**Species Complex**	**Systems Found**
H. vittatus	Vittatus complex	Congo, Lake Tanganyika, Zambezi, Okavango, Coastal systems below Lower Zambezi
H. goliath	-	Congo
H. forskahlii	-	Sanaga, Nile, Omo
H. brevis	-	Nile, Gambia
H. tanzaniae	-	Ruvu, Rufiji
A	-	Congo
B	Vittatus complex	Lake Tanganyika (Lufubu)
C	Vittatus complex	Congo, Zambian Congo
D	Vittatus complex	Zambian Congo

The *H. forskahlii* lineage is divided into two monophyletic groups (see Figures 2.3-2.5 and Table 2.3), one of samples from the Sanaga River in Cameroon (haplotypes 15 – 17) and the other of samples from the Nile River and Lake Chamo, part of the Omo system in Ethiopia (haplotypes 18 – 20). This divergence testifies to an internal structure within this lineage which reflects allopatry across an icthyofaunal province. This discovery reveals previously unsuspected phylogenetic complexity within *H. forskahlii*. A more detailed study of this Forskahlii complex will likely reveal interesting relationships between populations and also establish whether this mtDNA lineage is composed of any other distinct lineages that exhibit independent evolutionary trajectories. The *H. tanzaniae* lineage is comprised of a divergence, with strong support, between the samples gathered from the Ruvu River (haplotypes 33 and 34) and those obtained from the Kilombero River, part of the Rufiji-Ruaha system (haplotypes 29 – 32). Further sampling, particularly from the Ruvu where only two samples were obtained, would be necessary to further investigate this split.

Within *H. vittatus*, three well supported groups were recovered in the two Bayesian trees. These groups consisted of samples from the Congo and Lufubu Rivers (haplotypes 1-3, 5, 12 and 14), the Upper Zambezi and Okavango Delta (haplotypes 6, 7 and 10) and the Middle-Lower Zambezi and Shire Rivers and the coastal populations (haplotypes 8, 11 and 13). These groups reveal a signal of phylogeographic structuring that is associated with drainage system. This topic is explored in greater detail in Chapter 3.

As no complete molecular phylogeny of *Hydrocynus* has yet been published incorporating all described species, the phylogeny produced in this study significantly increases knowledge of the relationships between the species in this genus. The fact that ten unique evolutionary lineages were identified in all three mtDNA trees, while only five species are recognised in the current taxonomy of Hydrocynus, indicates that there is a greater diversity of tigerfish in Africa than previously thought. In particular, it is now known that lineages not associated with previously described species occur across the Congo basin.

It is noteworthy that all these lineages occur in sympatry with other lineages of *Hydrocynus* that represent described species. Therefore, these lineages either evolved in sympatry, possibly coming to occupy different ecological niches, or they were separated by barriers to their distribution such that the ancestral population evolved in allopatry and then subsequently dispersed to result in the sympatric distribution seen today. Future ecological and behavioural studies of the *Hydrocynus* genus would help to clarify if these lineages occupy different niches. Since this is beyond the scope of this study and currently recognised *Hydrocynus* species occur sympatrically with no evidence of their occupying different ecological niches, it is hypothesised here that these lineages are the result of allopatric speciation. A thorough phylogeographic study of these lineages should resolve their areas of origin. As this study was confined to an examination of mtDNA, the lineages identified represent maternal lines of descent. The results from nuclear sequence data in combination with morphological analysis will help to clarify whether these lineages in fact represent novel and undescribed species of *Hydrocynus*.

In the dated *cyt b* tree resolved by BEAST (Figure 2.6), the same ten unique evolutionary lineages observed in the other three trees were recovered with generally high support. However, *H. forskahlii* and *H. brevis* showed a poorly supported sister relationship in this tree which was not recovered in the other three trees. The divergence between these two species and the remaining taxa was also not well supported, so this branching arrangement and sister relationship is questionable. As the Birth-Death Speciation prior was used in the construction of this dated BEAST tree, the difference in tree topology could be a function of this algorithm. Due to the weak support of this branch, no confidence limits on the dating of this event are given, making the estimate of the date of the divergence between these two lineages highly questionable. Divergences dates and their confidence intervals (CI) are summarised in Table 2.4 and 2.5.

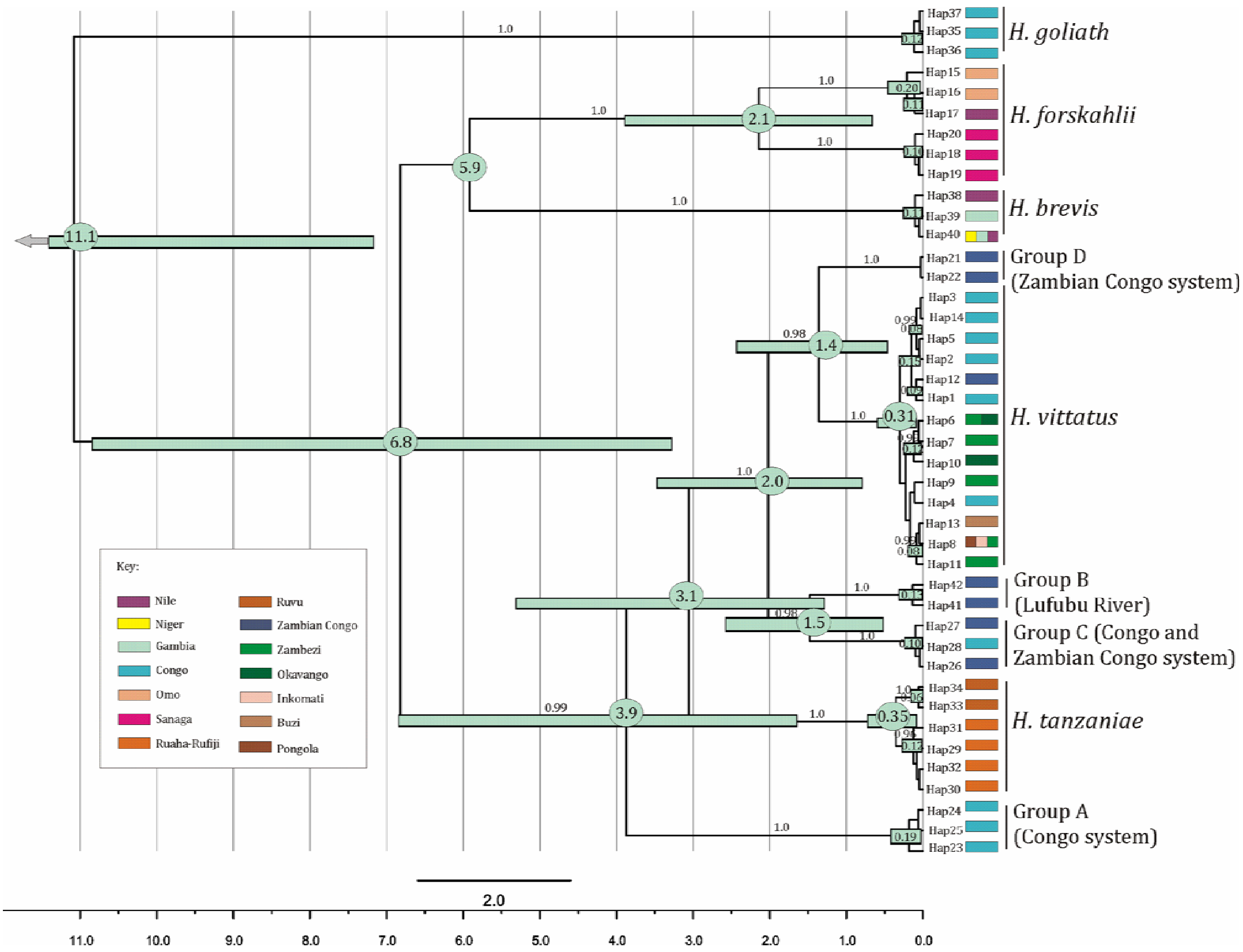


Figure 2.6: Dated Bayesian tree of the *cyt b* sequence data of *Hydrocynus* produced in BEAST, using the GTR parameters specified by Modeltest

Only posterior probabilities over 0.70 are shown. Error bars indicate the 95% confidence intervals on the node dates; numbers inside the bars are the dates estimated by BEAST; those inside ovals are those discussed in the text. The horizontal axis depicts time before present in Millions of Years. The colour key indicates the drainage system(s) represented in each haplotype.

Table 2.4: Summary table of the divergences between main lineages recovered in the dated BEAST tree including the 95% CI's

Divergence between		**Lower 95% CI**	**Estimate (in Ma)**	**Upper 95% CI**
H. goliath	All other taxa	7.1	11.1	15.5
H. brevis and *H. forskahlii*	A, *H. tanzaniae*, C, B, *H. vittatus* and D	3.2	6.8	10.8
H. brevis	*H. forskahlii*	-	5.9	-
A	*H. tanzaniae*, C, B, *H. vittatus* and D	1.6	3.9	6.9
H. tanzaniae	C, B, *H. vittatus* and D	1.3	3.1	5.3
C and B	*H. vittatus* and D	0.8	2.0	3.5
B	C	0.5	1.5	2.6
H. vittatus	D	0.5	1.4	2.4

Table 2.5: Summary table of the main divergences within lineages recovered in the dated BEAST tree including the 95% CI's

Main divergence within	**Lower 95% CI**	**Estimate (in Ma)**	**Upper 95% CI**
H. forskahlii	0.6	2.1	3.9
H. tanzaniae	0.1	0.35	0.7
H. vittatus	0.05	0.31	0.6

These estimated divergence dates have fairly large confidence intervals (e.g. approximately ±4 Million years on the divergence between *H. goliath* and the rest). This is most likely due to the use of only one calibration point and a universal estimate of substitution rates for teleost c*yt b*. However, despite these large confidence intervals, the estimated dates can be used as a starting point to discuss the evolutionary history of *Hydrocynus* since even approximate estimates can be used to test biological hypotheses (Bromham & Penny, 2003).

The age of the node separating *H. goliath* from all other *Hydrocynus* species is 11.1 Ma (CI: 15.5-7.1 Ma). The *H. brevis* and *H. forskahlii* clade diverged 6.8 Ma (CI: 10.8-3.2 Ma) but this is not well supported, as discussed above. The age range of this node is Late Neogene (late Miocene and Pliocene). Novel lineage Group A diverged approximately 3.9 Ma (CI: 6.9-1.6 Ma), also between the Miocene and Pliocene. The node separating *H. tanzaniae* and the Vittatus complex is dated in the Pliocene at 3.1 Ma (CI: 5.3-1.3 Ma). All divergence events of the lineages within the Vittatus complex fall within 3.5-0.5 Ma, the Plio-Pleistocene. During this time period several drainage rearrangements occurred in the Zambian Congo and the Zambezi systems, which are the areas where these groups are found today (see Table 1.1). The divergence between the two sister groups, Groups C and B and *H. vittatus* and Group D, is estimated at 2.0 Ma (CI: 3.5-0.8 Ma). Groups B and C diverged at 1.5 Ma (CI: 2.6-0.5 Ma) while *H. vittatus* and Group D split approximately 1.4 Ma (CI: 2.4-0.5 Ma). Divergences within some of the main lineages were also dated. Within *H. forskahlii*, the divergence between the two independent lineages, those from the Sanaga River in Cameroon (haplotypes 15 – 17) and those from the Nile River and Lake Chamo (haplotypes 18 – 20), dates to 2.1 Ma (CI: 3.9-0.6 Ma). The main divergence between the *H. tanzaniae* samples from the Ruvu River (haplotypes 33 and 34) and from the Rufiji River (haplotypes 29 – 32), dates to 350 ka (CI: 100-700 ka).

Four dated nodes were recovered with strong nodal support within *H. vittatus.* These reveal divergence events that all occurred relatively recently in the Pleistocene. The first, main divergence event, dated to 301 ka (CI: 50 - 600 ka), individuated a lineage comprising all but one of the Congo, Lufubu and Kwango haplotypes (haplotypes 1-3, 5, 12 and 14) from the Zambezi, the Shire, the coastal populations and the remaining Congo-Kwango haplotype. The grouping of the Congo, Lufubu and Kwango haplotypes was not well supported in this tree, in contrast with the results from the two undated Bayesian trees. The three other well-supported divergences constrain the origins of four Congo and Kwango haplotypes (haplotypes 3, 5, 12 and 14) to 150 ka (CI: 10 - 500 ka), all the Okavango and Upper Zambezi haplotypes (haplotypes 6, 7 and 10) to 120 ka (CI: 0 - 300 ka), and all but one of the Middle Zambezi, Shire and coastal population haplotypes (haplotypes 8, 11 and 13) to 80 ka (CI: 0 - 200 ka). The latter two groups are consistent with those well-supported groups recovered in the two undated Bayesian trees discussed above. These differences in support most likely reflect the differences in the statistical algorithms used by different programs and settings in reconstructing the relationships, as discussed above.

2.3.3 Nucleotide and Haplotype Diversity and Genetic Divergence between Lineages

The nucleotide and haplotype diversities for the *cyt b* data set (Table 2.6) show high haplotype diversities (Hd) and moderate to high nucleotide diversities of *H. vittatus*, *H. forskahlii*, *H. tanzaniae* and Group A, indicating stable populations with large, long term effective population sizes. Alternatively, these statistics could represent analysis of a mix of historically split populations since sampling of each lineage encompassed a large geographical area and often separate drainages. Group B also shows the abovementioned pattern but, since it consists only of two individuals, this could be a sampling artefact and more sampling is needed in order to carry out meaningful data analysis on this lineage. *H. goliath*, *H. brevis* and Group C have a moderate to high Hd but a low nucleotide diversity. This indicates rapid population growth from historically small effective populations. The high Hd of these lineages suggests active sequence divergence in the recent history of these lineages, which is characteristic of populations expanding in both size and distribution. Group D has very low nucleotide and haplotype diversities which indicate that this lineage has experienced a severe bottleneck in its recent history.

Table 2.6: Table of the descriptive statistics for the *cyt b* data set

	Number of individuals	**Number of haplotypes**	**Haplotype diversity (Hd)**	**Nucleotide diversity**
H. vittatus	32	14	0.8911 ±0.0312	0.0031 ±0.0019
H. goliath	8	3	0.6071 ±0.1640	0.0008 ±0.0008
H. forskahlii	8	6	0.9286 ±0.0844	0.0434 ±0.0241
H. brevis	9	3	0.5556 ±0.1653	0.0008 ±0.0007
A	5	3	0.7000 ±0.2184	0.0020 ±0.0016
D	9	2	0.2222 ±0.1662	0.0000 ±0.0000
C	7	3	0.5238 ±0.2086	0.0007 ±0.0007
H. tanzaniae	10	6	0.8444 ±0.1029	0.0046 ±0.0029
B	2	2	1.0000 ±0.5000	0.0037 ±0.0043

As none of the trees constructed were distance based, the Kimura 2-parameter genetic distances between groups and within groups were calculated to quantify the relative amount of differentiation. The genetic distances between groups (see Appendix 3.2 for data table) were very high overall (4.1% - 55.6%), indicating deep splits between the lineages depicted in the phylogenetic trees. *H. goliath* was the most distant from all other lineages, which is congruent with its basal position in all the phylogenies. However, its sequences differ from the rest by between 41.0 and 55.6%, which are indeed large distances. This large divergence points to saturation of the *cyt b* marker by nucleotide changes, likely at the third codon position since the amino acid sequences were extremely conserved, indicating selection against amino acid changes. This is common in fish *cyt b* evolution (Murphy & Collier, 1997) and it brings into question the validity of this marker for examining deep relationships between species (Farais *et al*., 2001). Confirmation of the topology found using this marker needs to be undertaken using nuclear markers appropriate for long evolutionary histories, as employed by Calcagnotto et al (2005). *H. forskahlii* and *H. brevis* show high genetic distances when compared with the other lineages. This is also congruent with their position in the phylogenetic tree. Members of the Vittatus complex were the least distant from each other.

The distances within lineages (see Appendix 3.3 for data table) were low except for within *H. forskahlii*, which had a value of 4.7%. This large distance points to genetic structuring within this species. This conclusion is supported by the phylogenetic trees which show a deep divergence between the two sister clades of *H. forskahlii* (see section 2.3.1 & 2.3.2).

2.3.4 Population Structure and Demographic History

The gsi values were all 1 ($p<0.005$), indicating that all the lineages shown in the phylogeny are monophyletic and on independent evolutionary trajectories. This finding from the genetic data confirms the classification of *Hydrocynus*, based on morphological evidence, which recognises five species. It also supports the independent evolutionary histories of the newly identified lineages initiated in the Plio-Pleistocene (Groups A, B, C and D).

The AMOVA showed that 97.4% of the variation was among populations and only 2.6% was within populations. This result of significantly more variation distributed among groups supports substantial genetic differentiation and monophyly between the different lineages, also seen in the gsi results and the phylogeny. All the population specific ΦST values were high (see Appendix 3.4 for data table), indicating a large amount of variation between groups. This is supported by the phylogenetic relationships recovered.

Tajima's *D* and Fu's *Fs* tests of selective neutrality examine the frequencies of mutations in order to detect deviations from the neutral model. Due to the small sample sizes (2-10 individuals) for all the lineages, excluding *H. vittatus*, it is not surprising that neither of these tests of selective neutrality were significant since the ability to detect deviations from the neutral model increases with sample size (Table 2.7). Only *H. vittatus* had a significantly negative value (-6.2153; $p < 0.005$) for Fu's *Fs*, which is more sensitive to recent population expansions (Fu, 1997). This result indicates an excess of low-frequency mutations which can be caused by demographic expansion or selection. However, as *cyt b* is a protein coding gene, it could be under selection to retain its functionality, resulting in the significantly negative Fu's *Fs*. This is supported by evidence of selection seen in the conservation of the amino acid sequences (see section 2.2.3).

Table 2.7: Table of lineage specific Tajima's *D* and Fu's *Fs* values

	Tajima's *D*	**Fu's *Fs***
H. vittatus	-1.264	-6.2153*
H. goliath	-0.4479	-0.4776
H. forskahlii	1.9978	2.9420
H. brevis	-0.5833	-0.5321
A	-1.0938	0.2764
D	0.0000	-0.2635
C	-1.2372	-0.9218
H. tanzaniae	0.0817	-0.3826
B	0.0000	1.0986

*: Significant values for $p < 0.005$

Since *H. vittatus* alone had a significant value for Fu's *Fs*, only its mismatch distribution was calculated (Figure 2.7). The 'raggedness' index was high (r =

0.25) and significant ($p < 0.01$), indicating a poor fit of the observed and expected mismatch distributions (Rogers & Harpending, 1992). The observed multimodal distribution, which is characteristic of a historically stable population, is supported by the high nucleotide and haplotype diversity in *H. vittatus* which indicates a stable population with a large, long term effective population size. However, this can also indicate a mix of historically split populations but the low level of genetic distance within *H. vittatus* does not support this.

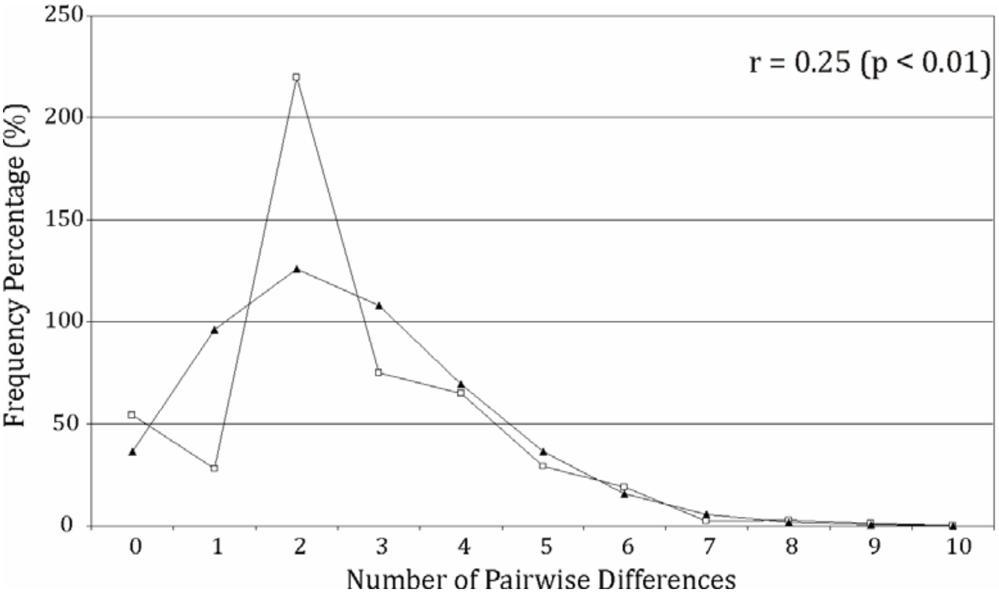


Figure 2.7: Mismatch distribution and 'raggedness' index (r) for the *cyt b* data of *H. vittatus*
Observed (□) and expected (▲) pairwise differences based on 10000 replicates.

2.4 Summary

The mtDNA phylogeny of *Hydrocynus* constructed using *cyt b* sequences is well resolved and supports the traditional taxonomy of this genus. An additional four novel lineages are uncovered in this study and deep internal divergences within *H. forskahlii* and *H. tanzaniae* are apparent. Based on the dated phylogeny, the initial speciation of *Hydrocynus* occurred in the late Neogene followed by a pulse of speciation in the Plio-Pleistocene, which resulted in the Vittatus complex. The population history of *H. vittatus* is relatively stable. A discussion of the findings in this chapter together with those of Chapter 3 will be provided in Chapter 4.

Chapter 3

Phylogeography of the *Hydrocynus vittatus* and the Vittatus complex

3.1 Introduction

In this chapter, the drainage history of the rivers of south-central Africa, from the Congo and Zambezi, south to the Pongola, is further explored through the phylogeographic structure of *H. vittatus* and the rest of the Vittatus complex. This complex consists of Group B (Lufubu River), Group C (Dja River, Lake Mweru and Lake Bangweulu) and Group D (Luapula River, Lake Mweru, Lake Bangweulu and the Chambeshi River). Sampling coverage of *H. vittatus* is represented by individuals from the Congo, Zambezi, Okavango, Zambian Congo and across the south-east Coastal systems south up to the Pongola. The findings of Chapter 2 provide the phylogenetic and biogeographical framework for this chapter which examines how the evolution of these tigerfish populations has interfaced with landscape evolution. The nomenclature of the lineages discovered in this study follows that assigned in Section 2.3.2 (see Table 2.3).

3.2 Materials and Methods

3.2.1 Sample Collection

Fin clips or muscle tissue were collected from 240 *Hydrocynus* specimens sampled from 12 drainage systems in the study area (see Figure 2.1, section 2.2.1). Appendix 1 lists all the samples, their haplotypes and localities. Samples were preserved in 96% (v/v) ethanol and stored at room temperature or 4°C until DNA was extracted.

3.2.2 DNA Isolation, Amplification and Sequencing

Depending on the quantity of tissue available for each sample, DNA extraction was performed using two different protocols. The chelex extraction protocol (Walsh *et al.,* 1991) was performed on samples represented by a sufficiently large piece of tissue, while when only a small piece of tissue was available, extraction was employed a QIAamp DNA Micro Kit (Qiagen) as per the manufacturer's instructions (see section 2.2.2 for details).

A combination of primers designed in this study and modified universal PCR primers for the mtDNA CR region were used to amplify this region. Primers used in this study were designed from an alignment of related species' sequences from Genbank (Benson *et al.*, 2005). The CR primers used in this study are detailed in Table 3.1.

Table 3.1: CR primers and their sequences (5' to 3')

Primer	**Sequence (5' to 3')**	**Source**
CRFISHA	CCGAAKATCGARGGTTAAASCCCTCC	Modified from Lee *et al.* (1995)
AHyc1F	AATGACTTAACTGATCCTGYXCTAG	Designed for this study
AHyc1seqF	AAGCRTCGGTCTTGTAATCCG	Designed for this study
CRFishE	CCTGAAGTAGGAACCAGATGCCAG	Modified from Lee *et al.* (1995)

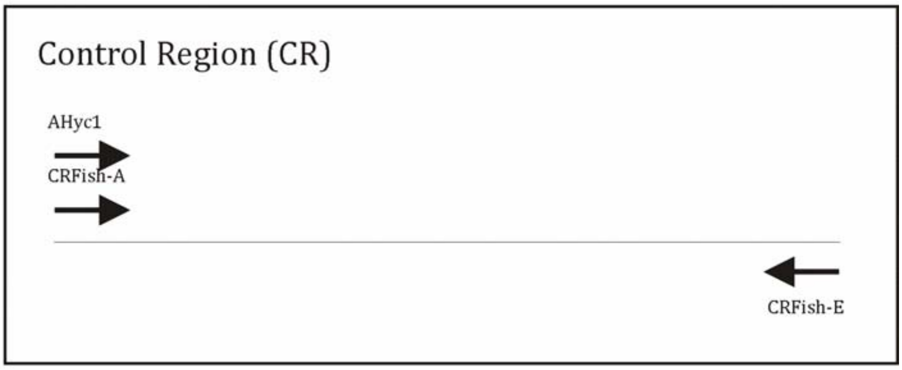


Figure 3.1: Schematic representation of the positions of primers for amplification of the CR of *Hydrocynus*

The primer pairs CRFish-A and CRFish-E or AHyc1 and CRFish-E were used for amplification (Figure 3.1). For both of these primer pairs, the PCR reaction mixture per sample (total volume 50µL) was composed of 0.5 units SuperTherm Taq (Southern Cross Biotech), 200µM dNTP's, 1x buffer, 2mM $MgCl_2$, 0.5µM of each primer, 1mg/µL bovine serum albumin (BSA) and 10-50ng template DNA. Approximately 20µL of mineral oil was added to each sample as an overlay. Reactions were carried out on a Programmable Thermal Controller-100 (MJ

Research, Inc.) with the following PCR conditions: 94°C for 5 min; 5 cycles of amplification composed of 94°C for 45 sec, 48°C for 45 sec and 72°C for 1 min; 30 cycles of amplification composed of 94°C for 45 sec, 50°C for 45 sec and 72°C for 1 min; and a final extension step of 72°C for 10 min.

PCR products were electrophoresed on 2% (v/v) agarose gels (1x TBE buffer), the target band was cut from the gel and cleaned up using a Wizard® SV Gel and PCR Clean-Up System column (Promega) and the purified DNA fragments were sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on an ABI 2700 thermocycler (Applied Biosystems). The sequencing reactions were run on an ABI 3100 sequencer (Applied Biosystems) (see section 2.1.2 for details).

3.2.3 Data Analyses

Sequence Editing and Alignment

DNA sequences were edited using Chromas 2.13 (Technelysium, 2006). The ClustalW feature (Thompson *et al*., 1994) in Bioedit 7.0.5.2 (Hall, 1999) was used to generate alignments. Alignments of *H. vittatus* sequences (131 sequences) as well as Vittatus complex sequences (204 sequences) were generated and adjusted by eye where necessary (see Appendix 4 for the 369 bp CR alignment). In both the CR alignments an indel (insertion or deletion) of between 36 - 206 bp was found. Indels are commonly found in the CR of fish (Billington & Hebert, 1991; Tang *et al.*, 2006). The indel in this data set was edited out to incorporate only the substitutions for further analyses since most analytical programmes' algorithms do not account for indels.

Phylogenetic Analysis

Paup 4.0b10 (Swofford, 1998) and ModelTest 3.06 (Posada & Crandall, 1998) were used to establish which model of nucleotide substitution fit the sequence data sets best as determined by the hierarchical likelihood ratio tests (hLRT).
Lists of unique sequences (haplotypes) for the CR data set were generated using Collapse 1.2 (Posada, 1999) (see Appendix 1). CR trees using the *H. vittatus* complex data set were produced using BEAST 1.4.8 package (Drummond &

Rambaut, 2007), MrBayes 3.1.1 (Huelsenbeck & Ronquist, 2001) and PAUP 4.0b10 (Swofford, 1998) as discussed in section 2.3.3. *H. tanzaniae* was selected as an outgroup for these analyses due to its sister relationship to the Vittatus complex as found in the phylogeny (see Chapter 2).

The program Splitstree 4.7 (Huson & Bryant, 2006) was used to construct Neighbor-net networks (Bryant & Moulton, 2004) for the Vittatus complex and the *H. vittatus* CR sequence data, using the selected model of nucleotide substitution. Split networks are useful to represent ambiguous signals in a data set by using parallel edges instead of single branches to represent the splits calculated from the data (Huson & Bryant, 2006). Ambiguous signals are often better represented in the form of a network when a large data set of intraspecific sequence data with low genetic distances between the sequences is being analysed (Bandelt *et al.*, 1999). The Neighbor-net algorithm, which is distance-based, is appropriate for use with large datasets, for example those from well-sampled phylogeographic studies (Bryant & Moulton, 2004).

Nucleotide and Haplotype Diversity and Genetic Divergence between Lineages

Genetic diversity was characterized by the haplotype and nucleotide diversities, estimated in Arlequin 3.11 (Excoffier *et al.*, 2005) and genetic distances within and between lineages, estimated in Mega 4.0 (Tamura *et al.*, 2007) using the Kimura 2-parameter model (Kimura, 1980). Rate heterogeneity among sites was corrected for with the gamma shape parameter.

Population Structure and Demographic History

Different *H. vittatus* populations were grouped for these analyses according to genetic divergence, sample size, location and their phylogenetic and network relationships. The reason for these groupings was the small size of several populations (see section 3.3.4 for further discussion).

The degree of common ancestry of groups of individuals on a rooted phylogenetic tree can be quantified using the genealogical sorting index (gsi)

(Cummings *et al.*, 2008). This statistic was used to quantify the common ancestry of CR sequences in a between-lineage context for the Vittatus complex and a between-population group context for *H. vittatus*. This analysis was performed as in section 2.2.3.

In order to assess the degree to which population structure reflects the current drainage patterns, an AMOVA (Excoffier *et al.*, 1992) was conducted and ΦST (Wright, 1951) was estimated. Tajima's *D* (Tajima, 1989) and Fu's *Fs* (Fu, 1996), which also indicate possible past population growth, were calculated. Mismatch distributions (Harpending, 1994) were used as an indicator of historical population size changes. The above three analyses were carried out in Arlequin 3.11 (Excoffier *et al.*, 2005). For details, see section 2.2.3.

Population Structure and Evolutionary History

Migrate 3.0 (Beerli & Felsenstein, 1999) was used to estimate the magnitudes and directions of inferred dispersal events (gene flow) between predefined population groups. This ability to reconstruct these events in the evolutionary history of lineages has potentially important applications to the understanding of how phylogeographic events relate to landscape topology and historical links between drainage systems. This set of analyses compares different scenarios of the evolutionary history of these tigerfish. Migrate uses signals in genetic sequence data to infer population genetic parameters in a maximum likelihood framework (Beerli, 2008). Populations were grouped by their region of origin, as well as on the results of the phylogenetic analyses. Migration was constrained to between geographically neighbouring populations since it was not considered possible for a fish to disperse between distant drainage basins without influencing the intermediate populations. These groupings and constraints are justified by the following reasons: to maximise the use of available computational resources, to minimise computational time (which, according to Bowie *et al.* (2006), can extend into weeks for analyses on even five populations), to best represent the possible migration scenarios within a freshwater environment and to overcome the problem of small sample sizes (Pfenninger & Posada, 2002) in certain populations. Preliminary analyses sampled the populations according to different structures in order to select the optimum grouping of populations for the main analyses. Five groups were analysed using Migrate in this study: a) Okavango; b) Upper Zambezi; c) Congo and Kwango; d) Middle and Lower Zambezi and Lower Shire; and e) the Coastal populations (Buzi, Save, Pongola, Inkomati and Usutu). The Lufubu population was excluded due to its distant geographical location in relation to the other populations. Unsampled populations can have an effect on Migrate results, the scale of which is dependent on the amount of the migration from these unsampled populations (Beerli, 2004).

Migrate calculates M, which represents the immigration into a population. This statistic reflects the incidence of gene flow and provides an index of immigration between any two populations within the sampled matrix. It is scaled to mutation rate (μ) and these statistics are specific to each Migrate analysis (Beerli & Felsenstein, 1999). Four independent runs were conducted to ensure convergence of the estimated parameters. F_{ST} values were used as starting values for population parameters of theta the mutation rate in each run. For every run, 10 short chains with a total of 10 000 000 generations and three long chains each with a total of 5 000 000 generations were both sampled every 100 steps. A burnin of 10 000 genealogies was discarded. A four chain adaptive heating scheme was used to improve searching of the data space. The results of the long chains were combined for the final estimates. Default values were implemented for the other settings.

3.3 Results and Discussion

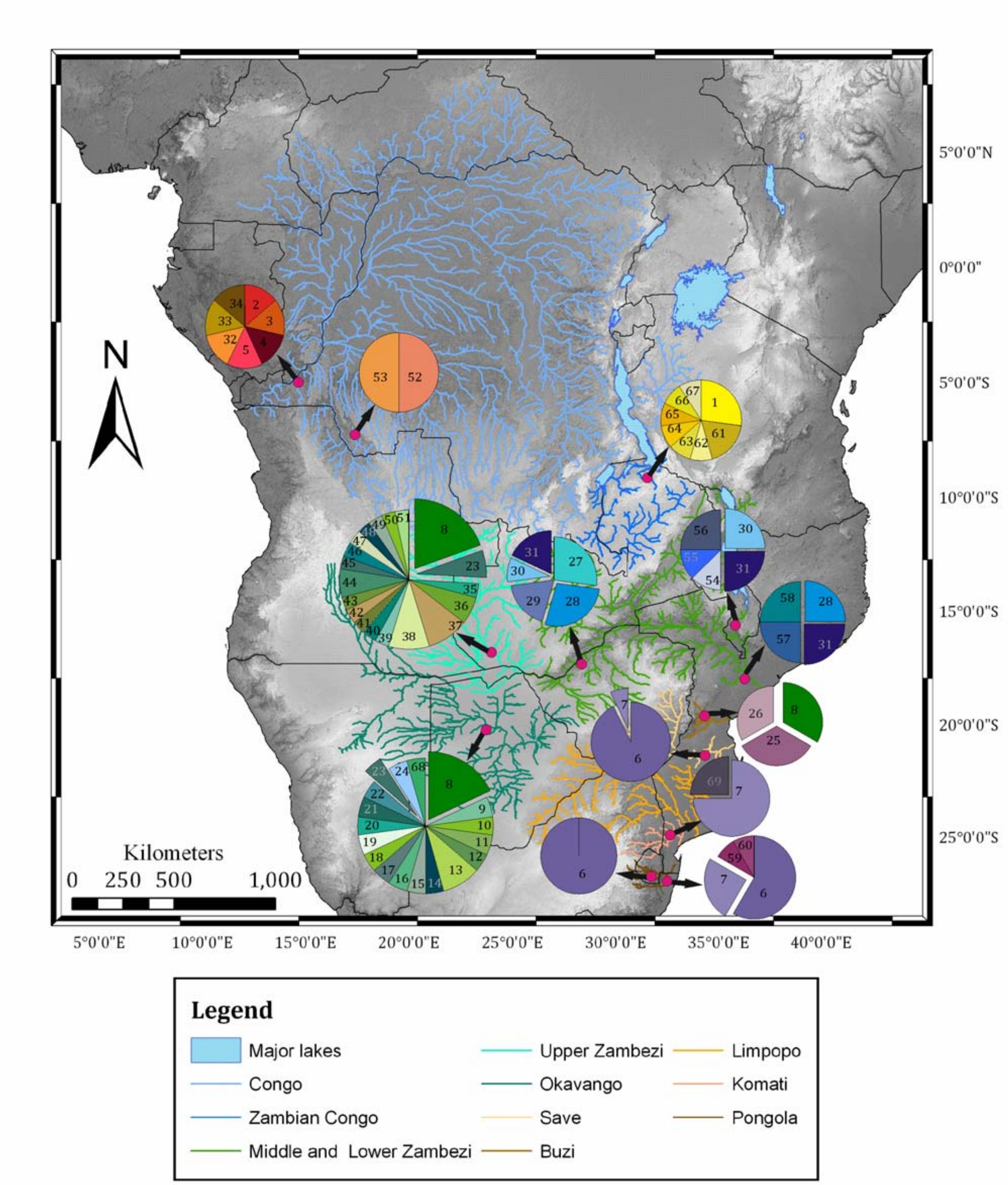
Analyses of the mitochondrial CR sequence data from the Vittatus complex and *H. vittatus*, within a phylogeographic framework, enable quantification of spatial and temporal patterns of genetic variation in this species complex. These results will be presented and discussed below in the context of tigerfish evolution, giving special attention to historical signatures that are interpreted to relate to respective landforms and events.

3.3.1 Data and Haplotypes Summary

For the CR Vittatus complex data set of 204 taxa, 117 haplotypes were recovered while for the CR *H. vittatus* only data set of 131 taxa, 69 haplotypes were recovered. Haplotype designations are listed in Appendix 1. *H. vittatus* haplotypes were mapped to sampling locations (Figure 3.2).

The Hasegawa, Kishino and Yano (HKY) model (Hasegawa *et al.*, 1985) plus Gamma (G) was selected for the Vittatus complex CR data (see Appendix 5.1 for data table). This model accounts for unequal base frequencies and a different rate of transitions (Ti) and transversions (Tv). This region was found to be A-T rich at 69%. A high percentage of A-T in the CR of other fish has also been found. For example, the mullet, *Mugil cephalus*, has 67% A-T (Rocha-Olivares *et al.*, 2005) and the Atlantic cod, *Gadus morhus*, 64% A-T (Johansen *et al.*, 1990). The Gamma shape parameter (α) accounts for rate heterogeneity among sites and an α of 0.3348 indicates extreme rate variation across the data. The TrN model (Tamura & Nei, 1993) plus G was selected for the *H. vittatus* CR data (see Appendix 5.1). This model accounts for unequal base frequencies and an excess of transitions. The CR was also found to be A-T rich (69%) in this analysis. An α of 0.2498 indicates extreme rate variation across the data.

Haplotypes shared between individuals from different locations are discussed below. *H. vittatus* haplotypes are the same for both data sets in this chapter (haplotypes 1 - 69). Haplotypes 6 and 7 are shared between *H. vittatus* from the coastal systems below the Zambezi River (Buzi, Save, Inkomati, Usutu and Pongola). This implies a relatively recent connection and/or gene flow between these systems. A relatively dry period occurred between 27 and 12 ka (LGM) which caused several rivers and lakes to dry up, while a more humid period is thought to have occurred approximately 12 to 8 ka and all lake levels were higher during that time (Roberts, 1975). *H. vittatus* was probably able to move from the Lower Zambezi in a founding event into the low-lying coastal rivers on the Mozambique coastal plain during a wet period as incidental connections formed when low-lying drainage divides on the Mozambique coastal plain were flooded (Roberts, 1975;



<u>Figure 3.2: Map of Africa showing the haplotypes from each collection locality for *H. vittatus*</u>
Map constructed using ArcMap 9.3 from the HydroSHEDS digital river database (Lehner *et al.*, 2008) and high-resolution elevation data from the Space Shuttle flight for NASA's SRTM (Jarvis *et al.*, 2006). Political boundaries are included for geographical context. Each haplotype is represented by a colour. Shared haplotypes are indicated by colour and the exploding slice in the respective pie chart; the number that represents that haplotype is listed in each slice (Appendix 1). (Due to the large number of haplotypes recovered in the Upper Zambezi and Okavango populations, the pie charts representing them have been enlarged to accommodate the haplotype numbers.)

Moore & Larkin, 2001), facilitating dispersal. This could have enabled the movement of *H. vittatus* between the coastal drainages, resulting in genetic admixture across this region.

Haplotype 23 is shared between the Okavango and the Upper Zambezi while haplotype 8 is shared between the Okavango, the Upper Zambezi and the Buzi Rivers. Intermittent connections between the Okavango and the Upper Zambezi systems during seasons of heavy rains (Bell-Cross, 1965a & 1972; Roberts, 1975) would account for this evidence of recent gene flow between these two systems. However, the haplotype shared between these two systems and the Buzi River is interesting because no Upper Zambezi or Okavango haplotypes are found in the Middle and Lower Zambezi or the Shire Rivers. However, the sample numbers from these rivers (12, 4 and 8, respectively) are low in relation to the size of the system. Therefore, it may be that Upper Zambezi-Okavango haplotypes do occur in this region but have not been sampled in this study. The hypothesised connection in the lower Cretaceous between the Palaeo-Limpopo, the Palaeo-Upper Zambezi and Palaeo-Okavango (Grove, 1969; Thomas & Shaw, 1991; Moore & Larkin, 2001) (see section 1.2.3) is not recent enough to account for this shared haplotype. This is because the CR marker, due to its highly variable nature, has an evolutionary window that extends only into the relatively recent evolutionary past. No other more recent connections between the Upper Zambezi and Okavango and the Buzi have been proposed in the literature. Interestingly, Bell-Cross (1973) observed a link between the Buzi and the Upper Zambezi fish fauna based on shared species that do not occur in the intervening systems. Although the shared haplotype recovered in this study between these systems and the Okavango is consistent with this link, no clear explanation for this shared fauna exists that can be applied to *H. vittatus*. Ecological factors are invoked by Bell-Cross (1973) to explain this anomaly but *H. vittatus* occurs throughout the systems between these two rivers, indicating that for this species ecological factors are unlikely to be the cause. It cannot be ruled out that the occurrence of this haplotype in the Buzi system may be the result of an unrecorded translocation of Upper Zambezi or Okavango *H. vittatus* into this system although it seems unlikely that this would have taken place over the great distance between these two systems when *H. vittatus* stocks are available in the systems adjacent to the Buzi.

Haplotypes 28, 30 and 31 were shared between the Middle and Lower Zambezi Rivers and the Shire River. As this region is considered a single biogeographical unit (Skelton, 1994), the occurrence of these shared haplotypes in the region is not surprising.

Haplotypes 88 and 96 are shared between individuals from Lakes Mweru and Bangweulu as well as those from the Chambeshi River. The presence of shared haplotypes across the Zambian Congo supports Bell-Cross's (1965a) hypothesis that the falls on the Luapula River do not restrict fish movement, which is supported by the phylogenetic data results. Overall, these data indicate a recent connection across the Zambian Congo system between Lakes Mweru and Bangweulu.

3.3.2 Phylogenetic and Phylogeographic Analysis

The Vittatus Complex

Phylogenetic trees of the Vittatus complex CR sequence dataset, recovered by all three programmes, exhibit strong congruence, which adds support to the recovered phylogenetic relationships. All these trees depict a consistent major branching order, with mostly high bootstrap and posterior probability values (Figures 3.3-3.5 depict the results from the BEAST, MrBayes and PAUP analyses). The same lineages within the Vittatus complex identified in the *cyt b* tree were recovered in the CR tree. This is to be expected since both genetic markers are located in the mitochondrion, which passes from generation to generation as a whole. For the 50% majority-rule consensus tree constructed in PAUP, 235 characters of the 369 character alignment were found to be constant and 32 of the variable characters were phylogenetically uninformative, leaving a total of 102 informative characters. The 50% majority-rule consensus tree has a Consistency Index of 0.385 and a Retention Index of 0.854.

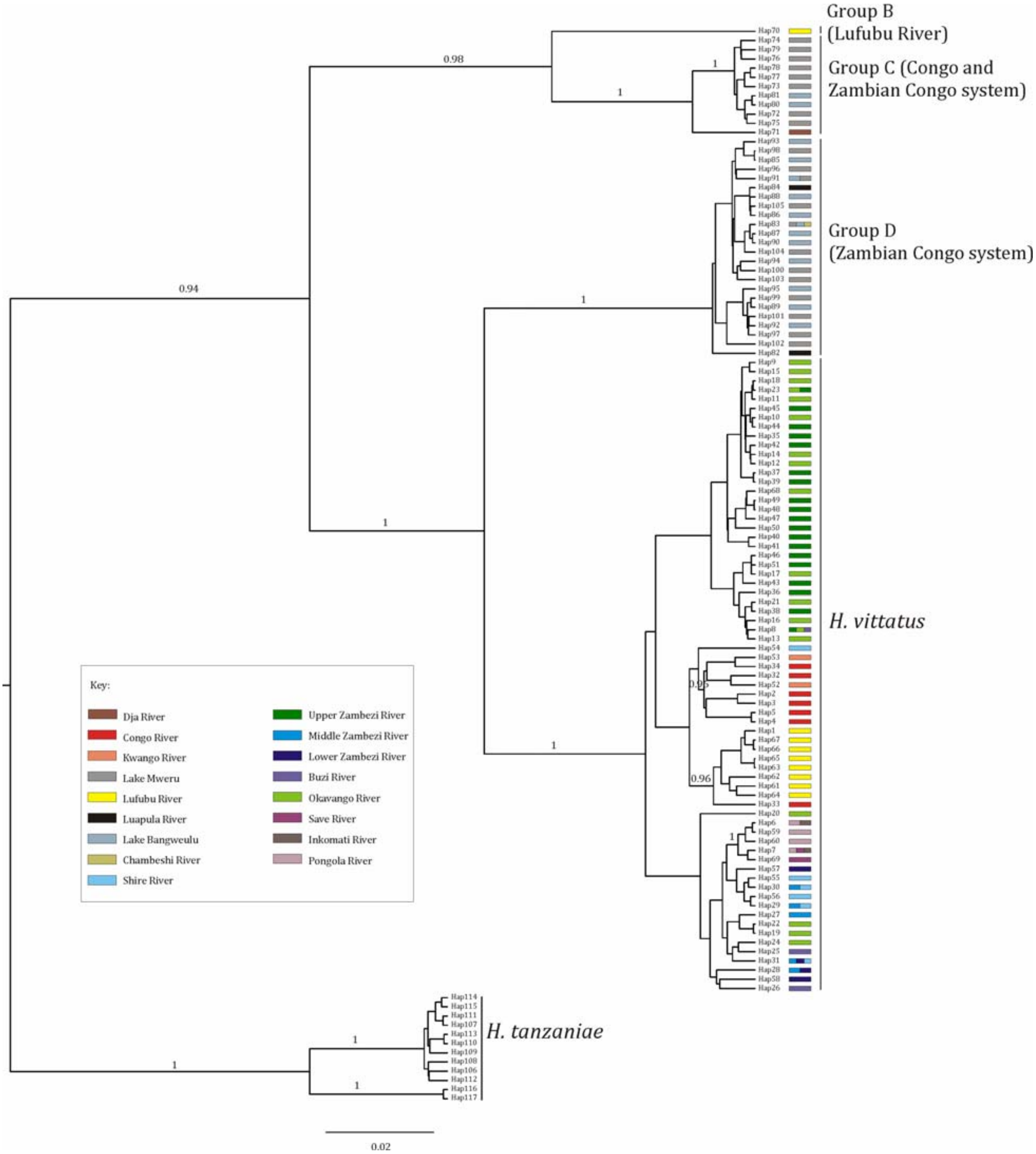


Figure 3.3: Bayesian tree of the CR Vittatus complex sequence data of *Hydrocynus* produced in BEAST, using the GTR parameters specified by Modeltest

Posterior probabilities over 0.70 are shown. Letters denote lineages not associated with previously described species as given in the text. The colour key indicates the river(s) represented in each haplotype.

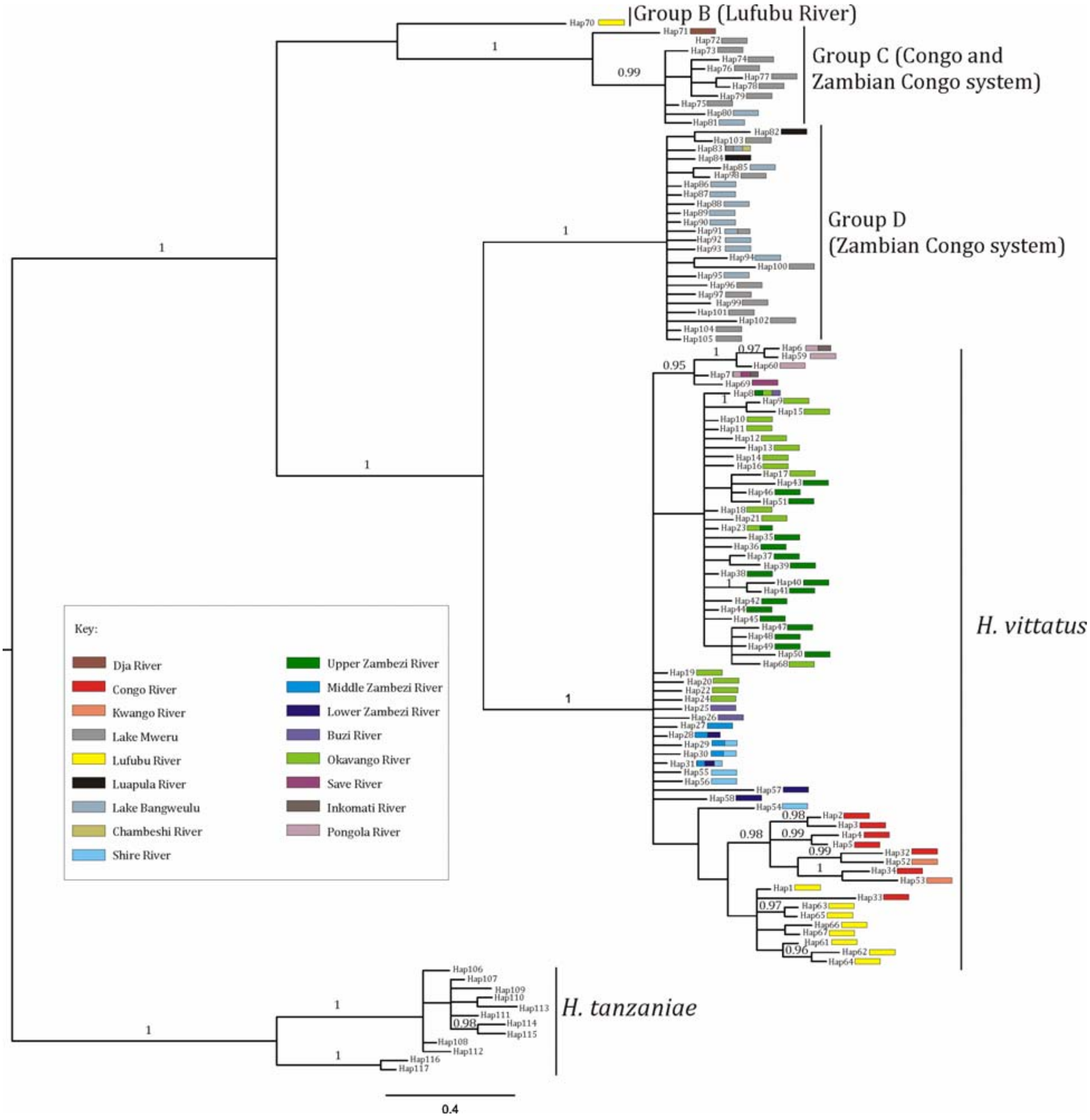


Figure 3.4: Bayesian tree of the CR Vittatus complex sequence data of *Hydrocynus* produced in MrBayes, using the GTR parameters specified by Modeltest

Posterior probabilities over 0.70 are shown. Letters denote lineages not associated with previously described species as given in the text. The colour key indicates the river(s) represented in each haplotype.

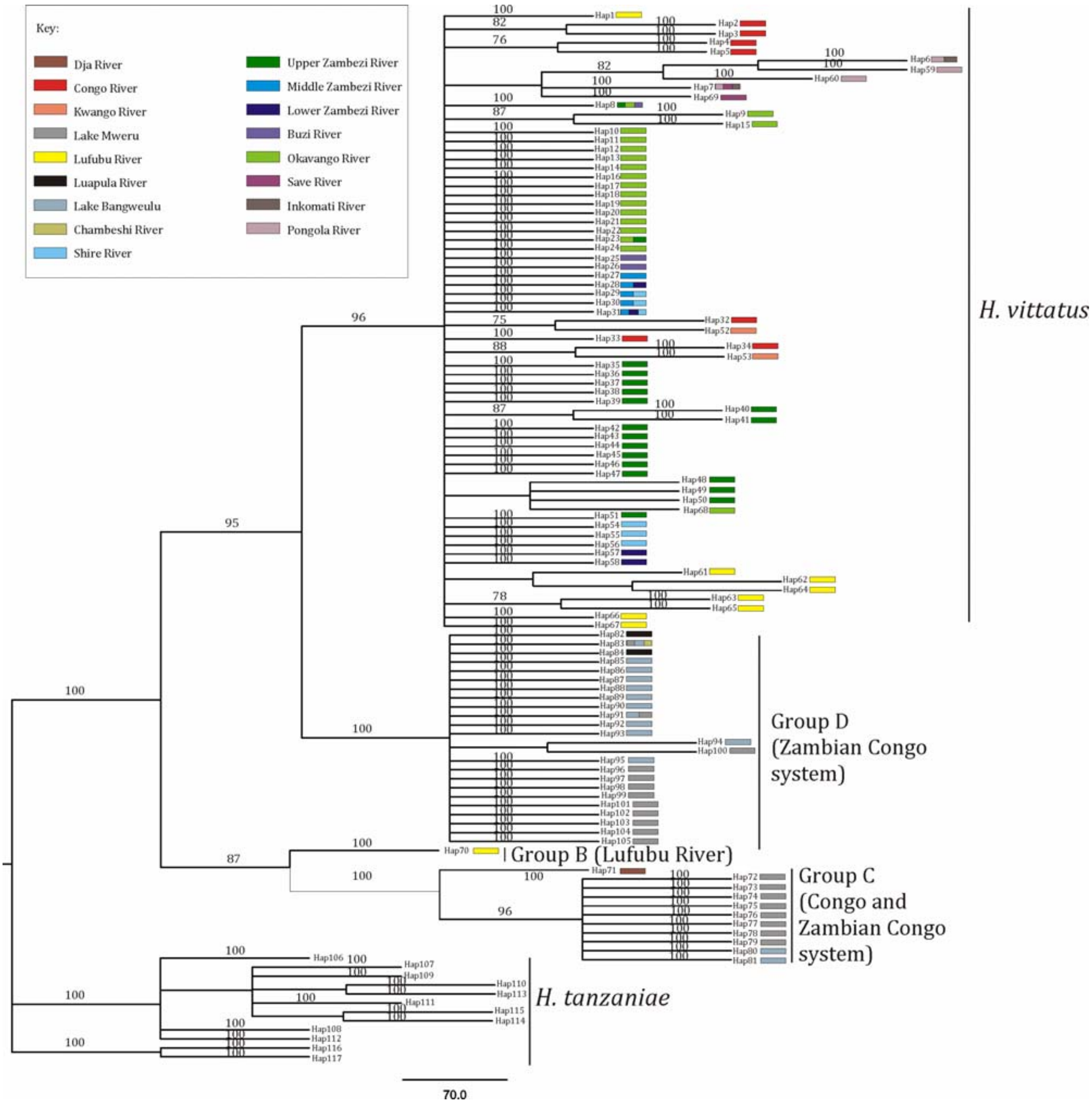
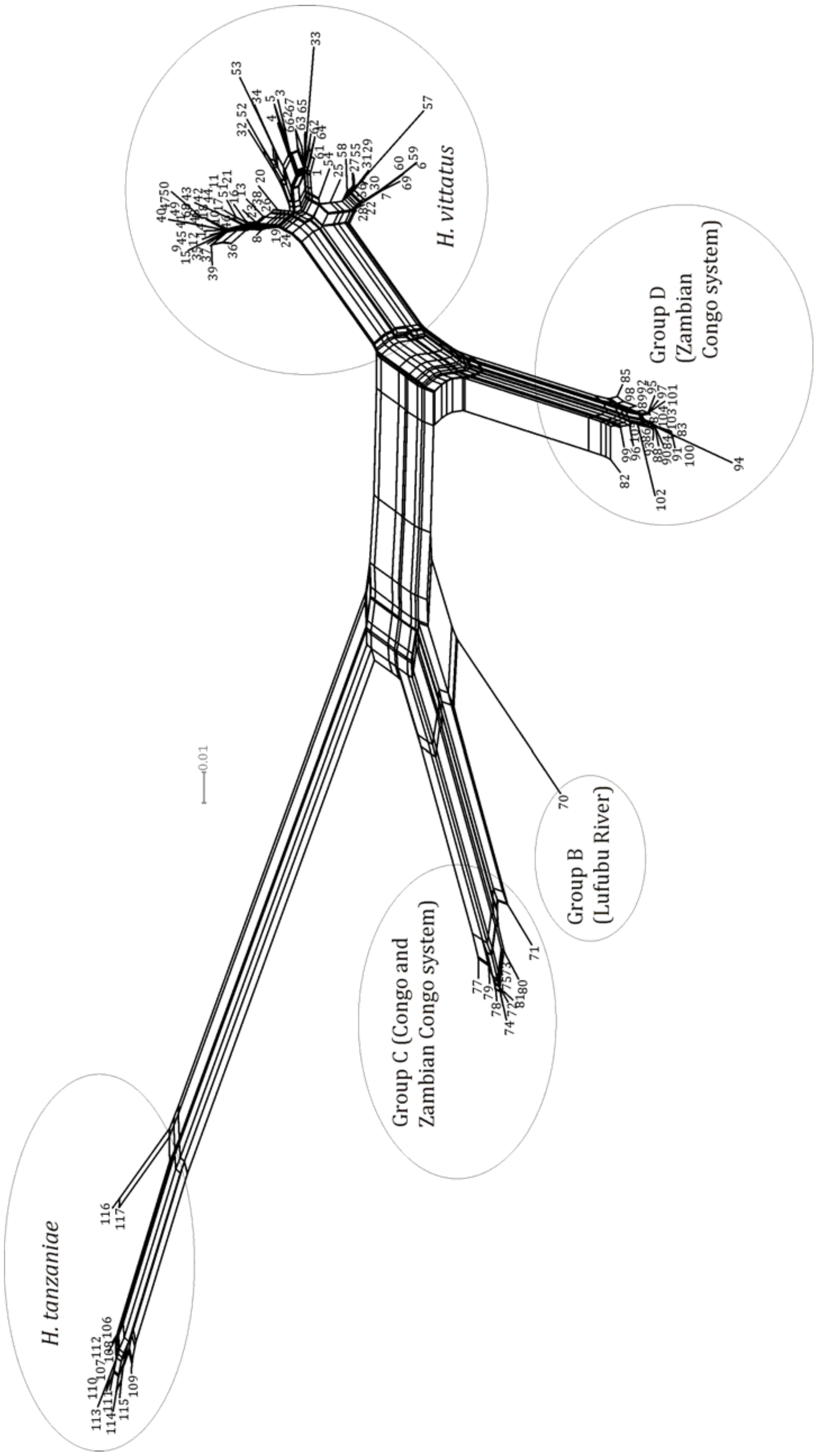


Figure 3.5: Maximum parsimony tree of the CR Vittatus complex sequence data of *Hydrocynus* produced in PAUP, using the GTR parameters specified by Modeltest

Bootstrap values over 50 are shown. Letters denote lineages not associated with previously described species as given in the text. The colour key indicates the river(s) represented in each haplotype.

The populations from the Ruvu River (haplotypes 116 and 117) and the Rufiji River (haplotypes 106 – 115) showed a divergence with regard to the *H. tanzaniae* lineage. The same divergence between these two systems was recovered in the *cyt b* tree. When evaluated in a geological timeframe, the sister relationship between the *H. tanzaniae* lineage from the Ruaha-Rufiji system (haplotypes 112 - 120) and an adjacent river, the Ruvu (haplotypes 121 and 122), indicates relatively recent separation between the two rivers. However, as there are only two samples from the Ruvu system, further sampling is necessary to evaluate this relationship more representatively. In Group C, haplotype 76 from the Dja River is basal to the rest of Group C which is composed of samples from Lakes Mweru and Bangweulu. The relationship is strongly supported and indicates that this ancestral event occurred somewhere within the main Congo channel between this western tributary and these lakes on the Kalahari Plateau. The divergence between Groups B and C indicates a historical separation between the Lufubu River and Lakes Mweru and Bangweulu. Both of these relationships were also recovered in the *cyt b* tree (see section 2.3.2). Within *H. vittatus*, three well supported groups were recovered in the BEAST tree and two of those were also found in the MrBayes tree. These groups consisted of samples from the Congo-Kwango (haplotypes 2-5, 32, 34, 52 and 53), the Lufubu-Congo (haplotypes 1, 33 and 61 - 66), although this group is not well supported in the MrBayes tree, and the coastal populations (haplotypes 6, 7, 59, 60 and 69).

The Neighbor-net network (Figure 3.6) of the Vittatus complex data reveals clear separations between the major evolutionary lineages, as were recovered by all three CR phylogenetic analyses and the *cyt b* tree. The same marked geographical structuring seen in the phylogenetic trees (see above) is revealed in the network for the *H. tanzaniae* lineage. In Group C, Haplotype 76 is from the Dja River, a tributary of the Sangha River of the Congo system. The rest of Group C is composed of samples from Lakes Mweru and Bangweulu. Three main groupings are seen in the *H. vittatus* lineage. These are the Congo-Kwango-Lufubu group, the Okavango-Upper Zambezi-Buzi group and the Middle-Lower-Zambezi-Shire-Coastal group. These groups will be further discussed in the following section on the Neighbor-net network of *H. vittatus* alone.



<u>Figure 3.6: A Neighbor-net network of haplotypes identified from Vittatus complex CR sequences, using the HKY Model of nucleotide substitution</u>

The numbers designated for each haplotypes are as listed in Appendix 1.

Overall, the strong geographical structuring in this CR dataset confirms the *cyt b* findings (see Chapter 2). The demographic history and evolution of *H. vittatus* will be focused on below.

Hydrocynus vittatus

The network of the *H. vittatus* CR sequence data (Figure 3.7) indicates the same major groupings in *H. vittatus* as was found in the Vittatus complex network. With distinct exceptions, there is a separation between the Upper Zambezi-Okavango populations and the Middle-Lower Zambezi-Shire populations. However, haplotype 22, an Okavango haplotype, groups with the Middle-Lower Zambezi-Shire group, while four Okavango haplotypes cluster centrally between the two groups. Haplotype 54, a Shire River haplotype, groups between the Middle-Lower Zambezi-Shire and the Congo-Kwango-Lufubu group. The samples from the Buzi also group centrally between the two groups. The Lower Congo and Kwango populations group together and exhibit a relatively close relationship with the Lufubu population, which flows into Lake Tanganyika. The coastal populations branch off from the Middle-Lower Zambezi-Shire populations.

In contrast to this network, phylogenetic analyses of the CR data set did not recover distinct phylogeographic structuring; they revealed no strong evidence for relationships between these groups. Only the Congo-Lufubu group shows a past relationship between the main Congo channel and the Lufubu River. The networks revealed exceptions to these groupings, which indicate past relationships between drainages. Several Okavango haplotypes show a degree of affinity between the Okavango and Middle Zambezi groups, possibly indicating an ancestral relationship between these two groups and beyond. Samples from the Buzi also group centrally between the two groups, supporting this relationship. The grouping of a Shire haplotype between the Middle-Lower Zambezi-Shire and the Congo-Kwango-Lufubu group could indicate an ancestral relationship of the Congo to the Middle Zambezi group. The Lufubu, Lower Congo and Kwango populations group together, indicating a relationship between *H. vittatus* across the Congo basin, which is supported by the findings of

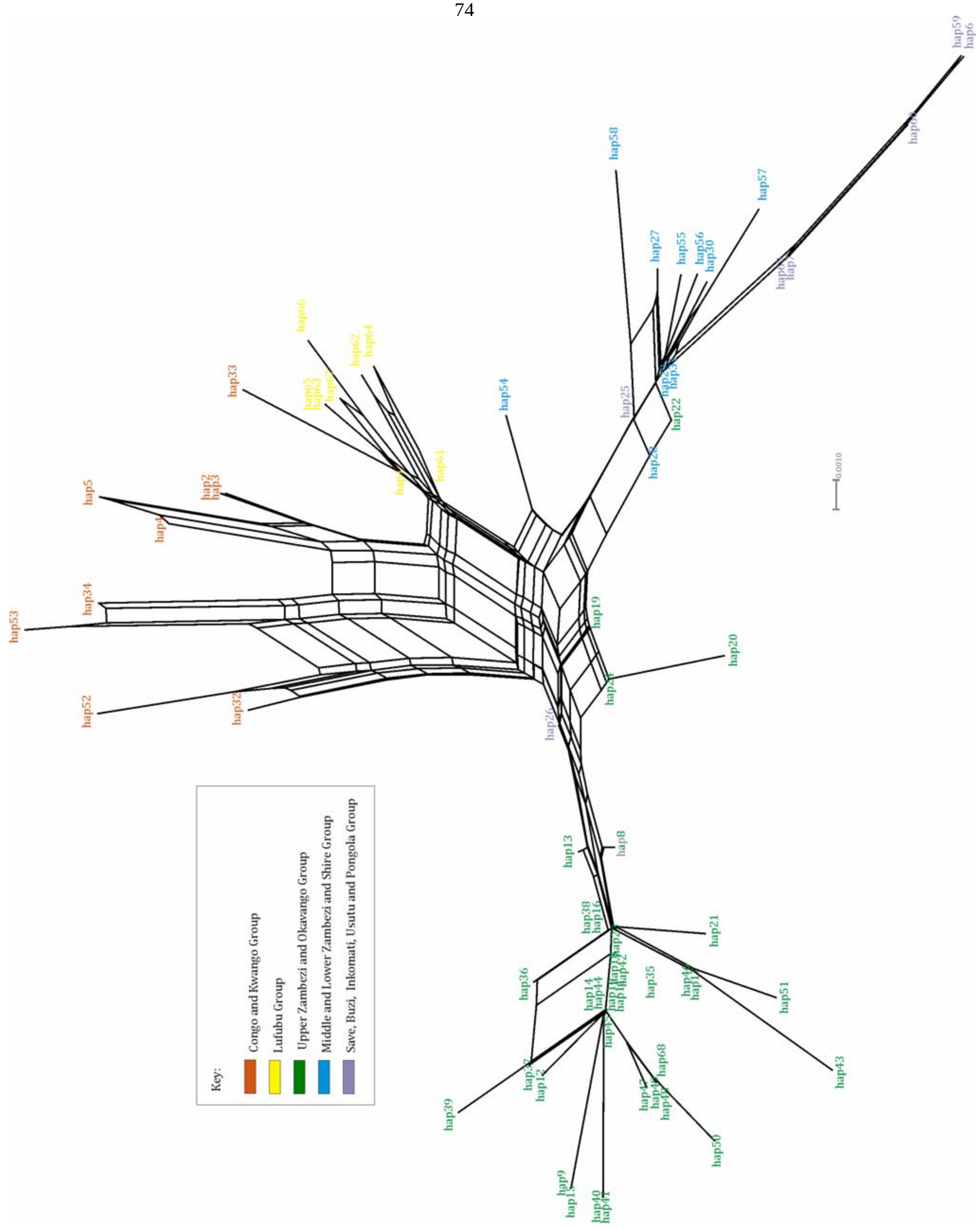


Figure 3.7: A Neighbor-net network of haplotypes identified from *H. vittatus* CR sequences using the HKY Model of nucleotide substitution

The colour key indicates the river(s) represented in each haplotype. The numbers designated for each haplotypes are listed as in Appendix 1.

the phylogenetic analyses. The coastal populations branch from the Middle-Lower Zambezi-Shire populations, indicating a founder event that populated the coastal systems from the Middle-Lower Zambezi-Shire lineage.

3.3.3 Nucleotide and Haplotype Diversity and Genetic Divergence between Lineages

The Vittatus Complex

For the CR marker data set (Table 3.2), the lineages *H. vittatus*, *H. tanzaniae* Group D and Group C are shown to have high nucleotide and haplotype diversities. This result indicates stable populations with large, long term effective population sizes or a mix of historically split populations. The same result was found for *H. vittatus* and *H. tanzaniae* in the analysis of the *cyt b* data set. Group B has no nucleotide and haplotype diversity which indicates that a lineage has passed through a severe bottleneck in its recent history. However, because the Group B lineage is represented by only two samples, these results could represent an artefact of insufficient sampling.

Table 3.2: Table of the descriptive statistics for the CR Vittatus complex data set

	Number of individuals	**Number of haplotypes**	**Haplotype diversity (Hd)**	**Nucleotide diversity**
H. vittatus	131	69	0.9592 ±0.0104	0.02316 ±0.0119
D	35	24	0.9580 ±0.0210	0.0070 ±0.0043
B	2	1	0.0000 ±0.0000	0.0000 ±0.0000
H. tanzaniae	22	12	0.8874 ±0.0540	0.0140 ±0.0079
C	14	11	0.9670 ±0.0366	0.0100 ±0.0061

Calculation of genetic distances provides revealing insight into the relative amount of differentiation among the Vittatus complex. Overall, the distances between Vittatus complex lineages are high, ranging from 8.7 - 26.9% (see Appendix 5.2 for the data table). *H. tanzaniae* is the most distant from all other lineages (19.7 – 26.9% genetic distance). A measure of the genetic distinctiveness of *H. tanzaniae* can be gauged against the maximum distance of 11% observed for the mtDNA CR in the genus *Tropheus*, a cichlid fish, endemic to

Lake Tanganyika (Baric *et al.*, 2003). Since it occupies the most basal position relative to the Vittatus complex in the *cyt b* tree and it was used as the outgroup in the CR analysis, this is an expected result. The distances within lineages range from moderate to low (0.8 – 1.0%) except for within *H. tanzaniae*, which has a value of 2.1% (see Appendix 5.3 for data table). This statistic indicates a level of structuring within this species, comprising the samples from two adjacent river systems that exhibit a sister relationship in both the cyt b phylogeny (see Chapter 2) and the CR tree.

Hydrocynus vittatus

Descriptive statistics were also calculated for the individual populations of *H. vittatus* for CR (Table 3.3). High haplotype and nucleotide diversity is found in the Congo, Kwango and Lower Zambezi populations, indicating large, stable population. High haplotype diversity but low nucleotide diversity is found in the Lufubu, Okavango, Upper Zambezi, Middle Zambezi, Lower Shire and Buzi populations. A historically small effective population that has grown rapidly in size would show this result. However, the low sample numbers of Buzi, Lower Zambezi and Kwango could affect these results and further sampling is needed to confirm these findings. The Save, Crocodile (a tributary of the Inkomati), Usutu (a tributary of the Pongola) and Pongola populations showed both low haplotype and nucleotide diversity. This very low nucleotide and haplotype diversity indicates that relatively recently these populations arose from a small founder population, which is consistent with a recent colonisation event (see section 3.3.1), or have passed through a severe bottleneck.

Table 3.3: Table of the descriptive statistics for the *H. vittatus* CR data set

Population	Number of samples	Number of haplotypes	Haplotype diversity (Hd)	Nucleotide diversity
Lufubu	11	8	0.9273 ±0.0665	0.0049 ±0.0034
Congo	7	7	1.0000 ±0.0764	0.0222 ±0.0134
Kwango	2	2	1.0000 ±0.5000	0.02406 ±0.0254
Okavango	22	18	0.9697 ±0.0278	0.0095 ±0.0056
Upper Zambezi	31	19	0.9484 ±0.0246	0.0071 ±0.0044
Middle Zambezi	11	5	0.8545 ±0.0657	0.0039 ±0.0029
Lower Zambezi	4	4	1.0000 ±0.1768	0.0153 ±0.0111
Lower Shire	8	5	0.8929 ±0.0858	0.0077 ±0.0052
Buzi	3	3	1.0000 ±0.2722	0.0108 ±0.0092
Save	4	2	0.5000 ±0.2652	0.0014 ±0.0017
Crocodile	14	2	0.1429 ±0.1188	0.0012 ±0.0013
Usutu	2	1	0.0000 ±0.0000	0.0000 ±0.0000
Pongola	12	4	0.6364 ±0.1277	0.0037 ±0.0028

Genetic distances between the different populations of *H. vittatus* range from 0.0 to 2.8% (see Appendix 5.4 for data table). This level of differentiation is similar to the distance range found at the intraspecific level in *Galaxius vulgaris* (0.0 – 2.6%), a freshwater fish found in New Zealand (Waters & Wallis, 2001). The Congo (1.4 – 2.5%) and Kwango (1.9 – 2.8%) populations are the most distant from all the other populations of *H. vittatus*. Because these Congo basin populations are also the most geographically distant from the rest of the populations, many unsampled populations occur between them and a sampling bias could account for the large distances obtained here. The Coastal populations (Save, Buzi, Inkomati, Usutu and Pongola) show low distances between them, as do the Middle and Lower Zambezi and the Shire populations. This indicates a recent colonisation event and/or mixing between the populations in the sampled area (see section 3.2.1). The distances within *H. vittatus* populations are low overall except for within the Congo and Kwango populations. These populations have values of 2.0 and 2.2%, respectively (see Appendix 5.5 for data table), indicating levels of geographical structuring within these two populations.

3.3.4 Population Structure and Demographic History

The Vittatus Complex

The gsi values for the five lineage groups were all 1 ($p < 0.005$). This indicates that all the lineages shown in the CR tree are monophyletic and on independent evolutionary trajectories. This finding is not unexpected, given that all phylogenetic analyses recovered a congruent topology comprised of evolutionary distinct lineages. These statistics do, however, support the independent evolution of the newly identified lineages (Groups B, C and D) also suggested by the *cyt b* gsi analysis .

The AMOVA showed that 91.07% of the variation is among lineages and only 8.93% is within lineages. All the lineage specific ΦST values are high at approximately 91.1% each (see Appendix 5.6 for data table). This indicates that significantly more variation is distributed within the groups than is shared among groups. These results point to substantial genetic differentiation between the different lineages, which concurs with the monophyletic estimates calculated with the gsi.

Tajima's *D* values are significant for Group D and C (Table 3.4). Fu's *Fs* is significant for all but Group B, for which it could not be calculated due to there being only one haplotype, and for *H. tanzaniae*. These significantly negative results indicate demographic expansion. As the CR is a non-coding region of the mtDNA genome it is unlikely to be under selection and so these results probably reflect a demographic expansion.

Table 3.4: Table of lineage specific Tajima's *D* and Fu's *Fs* values

	Tajima's *D*	**Fu's *Fs***
H. vittatus	-1.2864	-24.5722*
D	-2.1534*	-16.0752*
B	0.0000	N/A**
H. tanzaniae	-1.4314	-2.5994
C	-1.5981*	-5.1449*

*: Significant values ($p < 0.05$)

** Fu's *Fs* test cannot be computed if a sample comprises only one allele.

Only the mismatch distributions for *H. vittatus*, Group D and Group C were calculated (Figure 3.8a - c) since these are the only lineages with significant values for either or both of the neutrality tests. For all three lineages, a multimodal distribution of the observed mismatch was found, which is characteristic of a historically stable population. However, the 'raggedness' indices were low but not significant, indicating that an expansion event cannot be ruled out. Historical population stability is supported by the haplotype and nucleotide diversity analysis (see section 3.3.3).

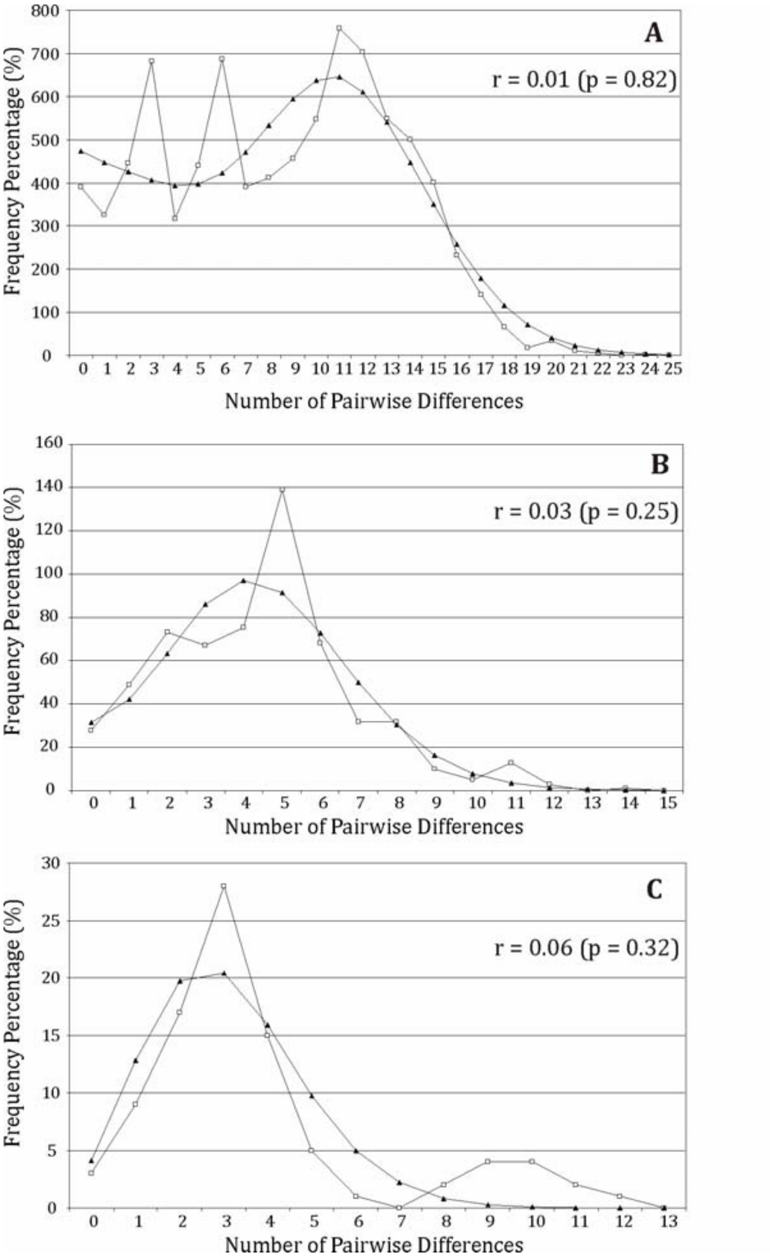


Figure 3.8: Mismatch distributions and 'raggedness' indices (r) for the CR data of *H. vittatus* (a), Group D (b) and Group C (c)

Observed (□) and expected (▲) pairwise differences based on 1000 replicates.

Hydrocynus vittatus

This section further explores the patterns of variation within and among thirteen populations of *H. vittatus* which here are grouped to facilitate statistical analyses of demographic histories. These populations were grouped based on the criteria of genetic divergence, sample size, location and structuring revealed in the phylogenetic trees and network. The coastal populations (Save, Usutu and Pongola) were grouped together due to the small divergence values between their populations and their relationship in the network. The Buzi population was included in this group since it is in the same geographic area. The Middle Zambezi, Lower Zambezi and the Shire were grouped together as they have low levels of genetic diversity between them and they group together in the networks. The Congo and Kwango populations grouped together in the network as well as the trees. The rest were left ungrouped.

For the analysis in which the different *H. vittatus* populations were analysed separately, the gsi values range from 0.1584 to 0.9044 (see Appendix 5.7 for data table). The low and moderate gsi values for all populations but Lufubu indicates that all these populations share a recent history with other populations through either recent admixture or incomplete lineage sorting. The Lufubu has a gsi value of 0.9044, which indicates an almost entirely independent evolutionary trajectory. In an effort to identify the populations that share a history, another run with the populations grouped was conducted (Table 3.5). These gsi values range from 0.3726 to 0.9044 and show shared histories between the Congo and Kwango, the Middle Zambezi-Lower Zambezi-Shire, and the coastal populations. The gsi of the Okavango population is noteworthy as it has a gsi on average that is almost half that of the other populations. An additional run combining the Okavango and Upper Zambezi indicated their shared history.

Table 3.5: Gsi values calculated to quantify the degree of shared evolutionary history between population groups of *H. vittatus*
P-values calculated based on a permutation test (10 000 replicates)

Population	GSI value	P-value
Lufubu	0.9044	< 0.001
Congo-Kwango	0.7918	< 0.001
Upper Zambezi	0.6089	< 0.001
Okavango	0.3726	< 0.001
Middle Zambezi-Lower Zambezi-Shire	0.7212	< 0.001
Coastal	0.7486	< 0.001

The population group AMOVA reveals that 70.99% of the variation is among population groups and that only 29.01% is within these groups. All the grouped population specific ΦST values are moderately high (see Appendix 5.8 for data table) with values between 68.1 and 71.1%. This indicates that a large amount of variation is distributed within the groups but an amount is also shared among groups, supporting a mixed history between the different population groups, which in line with the gsi findings.

Tajima's *D* values were significantly negative for the Okavango and Middle Zambezi-Lower Zambezi-Shire groups (Table 3.6). Fu's *Fs* was significant for the Lufubu, Congo-Kwango, Upper Zambezi and Okavango groups. These significantly negative results indicate demographic expansion.

Table 3.6: Table of *H. vittatus* population group specific Tajima's *D* and Fu's *Fs* values

Population	Tajima's *D*	Fu's *Fs*
Lufubu	-0.3748	-2.4510*
Congo-Kwango	-0.7985	-3.4183*
Upper Zambezi	-1.5215	-14.9645*
Okavango	-1.8005*	-12.7777*
Middle Zambezi-Lower Zambezi-Shire	-1.7321*	-2.3718
Coastal	-1.3512	0.2429

*: Significant values ($p < 0.05$)

The mismatch distributions for Okavango, Upper Zambezi, Lufubu, Congo-Kwango and Middle Zambezi-Lower Zambezi-Shire population groups were calculated (Figure 3.9a - e) as these lineages have significant values for the above neutrality tests. For the Okavango and Congo-Kwango population groups (Figure 3.9a & b), a multimodal distribution of the observed mismatch was found, which is characteristic of a historically stable population, but, just as discussed above for the Vittatus complex lineages, the 'raggedness' indices were low and not significant, indicating that an expansion event is possible. Lufubu and Upper Zambezi population groups (Figure 3.9c & d), a unimodal distribution of the observed mismatch was found which is characteristic of an expanding population. The 'raggedness' indices were low but not significant, supporting a possible expansion event. For the Middle Zambezi-Lower Zambezi-Shire population group (Figure 3.9e), a multimodal distribution of the observed mismatch was found which is characteristic of a historically stable population. The 'raggedness' index was high and significant, supporting a stable population, which is in contrast with the Tajima's D results. However, Fu's *Fs* was not significantly negative for this population group and it is this statistic which is the most sensitive to population expansions (Fu, 1997).

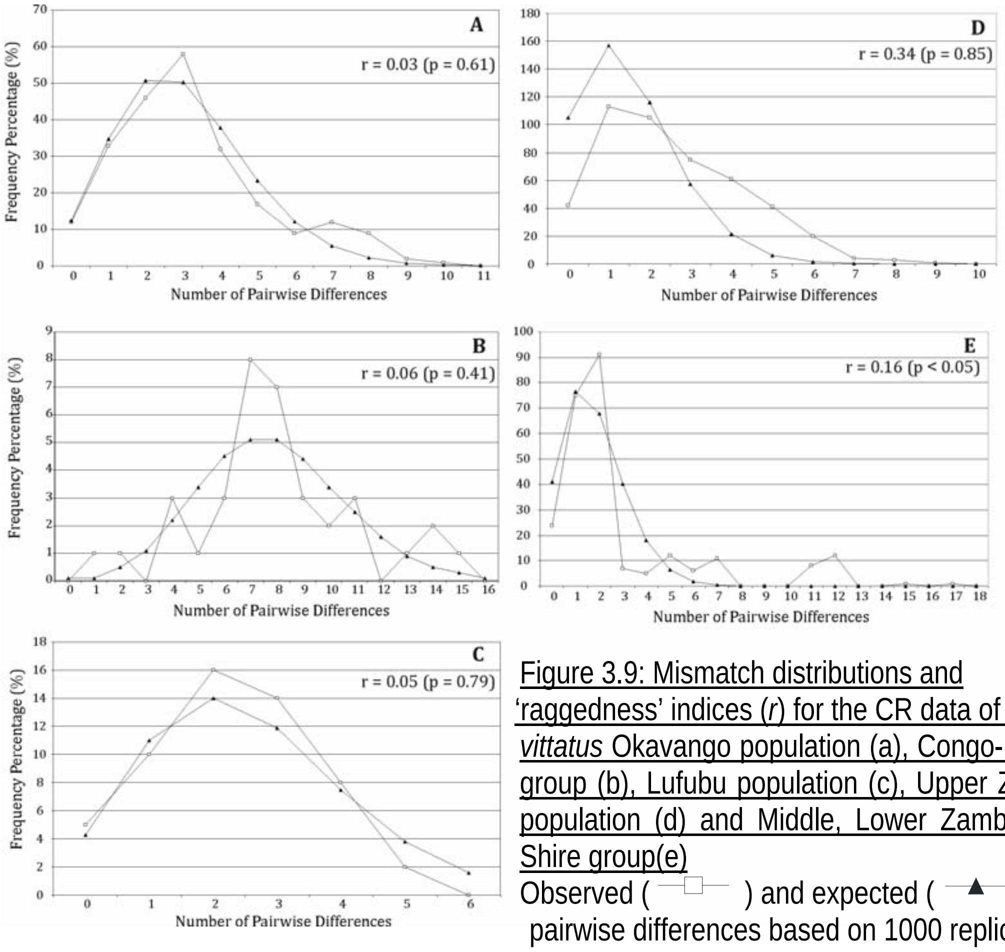


Figure 3.9: Mismatch distributions and 'raggedness' indices (*r*) for the CR data of the *H. vittatus* Okavango population (a), Congo-Kwango group (b), Lufubu population (c), Upper Zambezi population (d) and Middle, Lower Zambezi and Shire group(e)
Observed (—□—) and expected (—▲—) pairwise differences based on 1000 replicates.

3.3.5 Population Structure and Evolutionary History

The Migrate analysis performed on the five major *H. vittatus* population groups (Okavango, Middle Zambezi-Lower Zambezi-Shire, Congo-Kwango, Upper Zambezi and Coastal) indicates that all gene flow originated in the Upper Zambezi. Viewed in the context of the high degree of sequence diversity across populations, with respect to drainage basins, these Migrate statistics are interpreted as signals of past dispersal events. Each of these events has seeded populations of *H. vittatus* in the Congo, Okavango, and Middle and Lower Zambezi, which in turn seeded the coastal populations. Therefore each Migrate statistic can be interpreted as a proxy of the event that founded these populations, which, with the exception of the Okavango population, subsequently became isolated.

The ancestral Upper Zambezi population was the source for the Congo and Kwango, the Okavango and the Middle and Lower Zambezi and Shire. Immigrants into the coastal populations most likely dispersed from the Middle and Lower Zambezi and Shire (Figure 3.10). This suggests that the Upper Zambezi constitutes the source population of migrants of *H. vittatus* that seeded the analysed populations elsewhere in south-central Africa. The greatest signature of historical dispersal is from the Upper Zambezi into the Okavango. As these systems were connected historically (see section 1.3) and are intermittently connected even today (Bell-Cross, 1965a & 1972; Roberts, 1975), this finding is indeed expected. However, the lack of dispersal in the opposite direction (Okavango into Upper Zambezi) implies that the connections only allowed *H. vittatus* movement in a single direction or that the Upper Zambezi is the initial founding population and the polymorphisms shared between them elucidate the findings of Migrate analyses (see above). Immigrants into the coastal populations have dispersed from the Middle and Lower Zambezi and Shire in moderate numbers. This is supported by the network relationship between these two population groups, which show the Middle Zambezi group as the parent population. Additional coalescent analyses using programmes such as MDIV (Nielsen & Wakeley, 2001) and IM (Hey & Nielsen, 2004) can test this interpretation of historical events. This is because these two programs simultaneously quantify an index of recent migration and an estimate of divergence time between compared lineages (Hey & Nielsen, 2004).

It should be noted that due to insufficient sampling, it is probable that several populations are not represented in the Migrate analyses, which can affect reconstructions of gene flow (Beerli, 2004). In order to ensure a full picture of *H. vittatus* dispersal, more samples need to be collected. This sampling deficiency particularly applies to the Congo system. Moreover, the population structuring shown here must be interpreted with caution due to the difference in sample sizes for the different lineages.

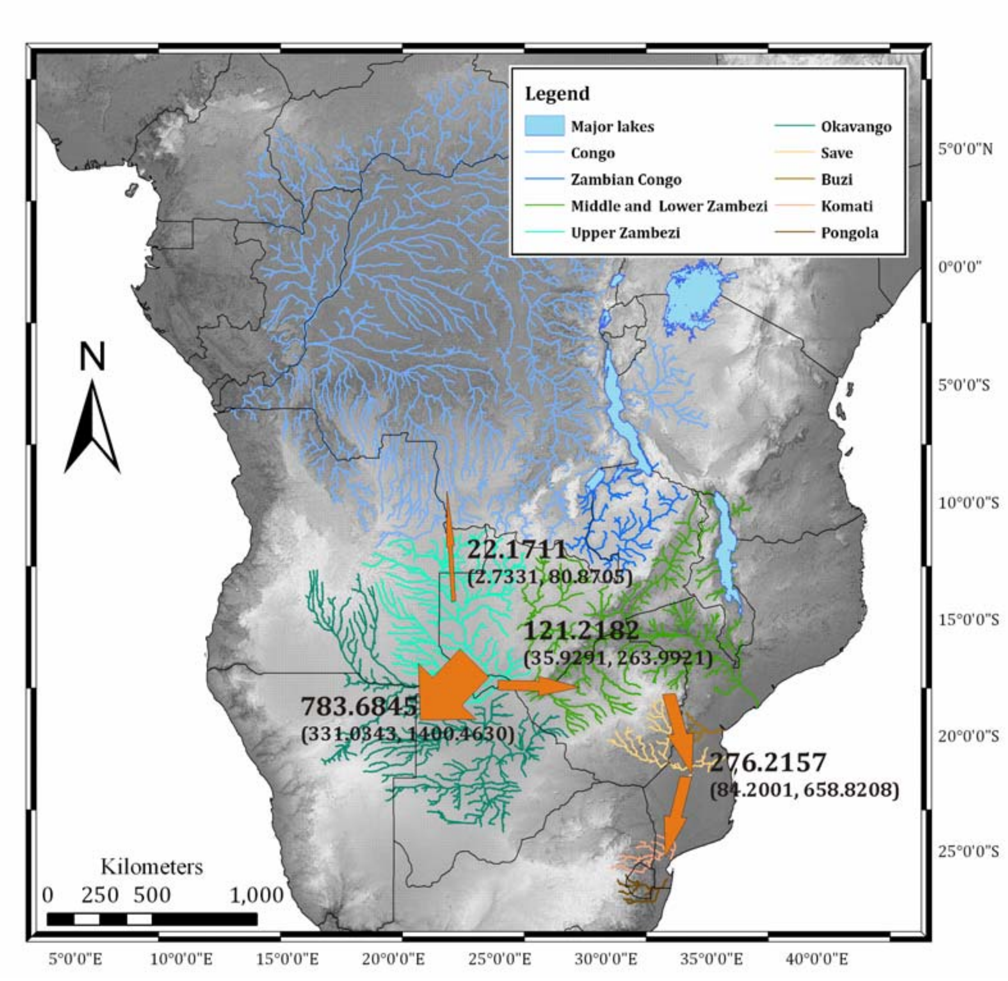


Figure 3.10: Map of south-central Africa showing the gene flow (M) between population groups of *H. vittatus* as calculated in the Migrate analysis

Map constructed using ArcMap 9.3 from the HydroSHEDS digital river database (Lehner *et al.*, 2008) and high-resolution elevation data from the Space Shuttle flight for NASA's SRTM (Jarvis *et al.*, 2006). Political boundaries are included for geographical context. Arrows and numbers represent the direction and amount of gene flow between population groups. The numbers in brackets represent the 95% confidence interval around each estimate.

3.4 Summary

The phylogenetic trees and the networks of the CR data agree with the *cyt b* results presented in the previous chapter (Chapter 2), supporting the discrete lineages labelled the Vittatus complex in this study. The independent histories of these lineages are supported by several of the analyses, such as the high gsi and ΦST values. In addition, the levels of both the nucleotide and haplotype diversity between the lineages in the Vittatus complex as well as between the *H. vittatus* populations are mostly relatively high. Strong geographical structuring is observed across the Zambian Congo system (Groups B - D) and within *H. vittatus*. The Migrate results indicate that within *H. vittatus* the Upper Zambezi population is likely the source that has seeded all other populations. A combined discussion of the above and the findings of Chapter 2 will be presented in Chapter 4.

Chapter 4

General Synthesis and Conclusion

4.1. Introduction

Phylogenetic analysis of sequence data and gene trees of the mitochondrial *cyt b* marker from *Hydrocynus*, together with phylogeographic analysis of sequence data from the non-coding CR marker from the Vittatus complex, have been discussed in the previous chapters (see Chapters 2 and 3). The aim of these analyses is to reconstruct the relationships between the different *Hydrocynus* lineages and to describe spatial and temporal patterns of genetic variation in the species. This historical reconstruction is then applied to test previously proposed hypotheses of drainage evolution with a focus on south-central Africa (see section 1.2.3). In this chapter, a synthesis of the evolution of *Hydrocynus* as discussed in Chapters 2 and 3 will be presented. This account will be followed by a discussion of the drainage hypothesis based on the evolutionary relationships of *Hydrocynus* lineages derived from the phylogenetic and phylogeographic data presented in Chapters 2 and 3. Finally, future opportunities for research identified by this study will be outlined.

4.2 General Synthesis

The evolution of *Hydrocynus* will be synthesised by focusing on the results discussed in Chapters 2 and 3 that involve the same lineages.

4.2.1 Haplotype Summary

Shared haplotypes between different geographic regions indicate recent gene flow between the two sampled areas (see sections 2.3.1 and 3.3.1). Haplotypes of both *cyt b* and CR genetic regions are shared across the Zambian Congo. This is particularly evident between Lakes Mweru and Bangweulu, both of which were well sampled. These shared haplotypes support Bell-Cross's (1965a) hypothesis that the falls on the Luapula River have not served as a barrier to fish movement in recent evolutionary times. The intermittent connection between the Okavango and Upper Zambezi systems (Bell-Cross, 1965a & 1972; Roberts, 1975) is likely the reason for shared haplotypes between these two systems. Shared haplotypes for both genetic markers are evident across the sampled coastal drainages south of the Lower Zambezi. In addition, for the *cyt b* marker, a shared haplotype is evident between the coastal populations and the Middle Zambezi. This is consistent with evolutionarily recent colonisation of the coastal systems from the Middle Zambezi region, possibly after the LGM (approximately 12 - 8 ka) (Roberts, 1975).

4.2.2 Phylogenetic and Phylogeographic Analysis

In terms of the phylogenetic trees (see sections 2.3.2 and 3.3.2), all three analysis programs produced trees with consistent major branching orders which exhibit a strong topological congruence for both genetic markers. This, along with high bootstraps and posterior probabilities, lends support to the reconstructed phylogenetic relationships within *Hydrocynus*, focusing on the Vittatus complex. Four lineages not associated with previously described species were identified, three of which were included in the Vittatus complex data set (no. 2 - 4): 1) Group A was composed of samples from the Congo basin; 2) Group B consisted of two Lufubu River samples 3) Group C was composed of samples from the Dja River, Lake Mweru and Lake Bangweulu; and 4) Group D was composed of samples from the Luapula River, Lake Mweru, Lake Bangweulu and the Chambeshi River. Internal structuring was observed within *H. forskahlii*, *H. tanzaniae* and, to some extent, within *H. vittatus*. The divergence dates calculated using the *cyt b* data set will be discussed in section 4.3 below in the context of a drainage hypothesis.

The phylogenetic analysis and the construction of a network of relationships between the Vittatus complex CR data discovered the same structure as recovered for those lineages in the phylogenetic trees (see section 3.3.2). Limited geographical structuring of *H. vittatus* populations was found in the phylogenetic trees but network relationships revealed a greater degree of structuring by river system. Possible relationships between the Okavango and the Middle Zambezi, continuing south, and the Congo and Middle Zambezi group, respectively, were identified. The Lufubu, Lower Congo and Kwango populations group together, indicating a historical relationship between *H. vittatus* across the Congo basin, which is supported by the findings of the phylogenetic analyses. A founder event that populated the coastal systems from the Middle-Lower Zambezi-Shire lineage is also indicated.

4.2.3 Nucleotide and Haplotype Diversity and Genetic Divergence between Lineages

For the CR lineage data set (see section 3.3.3), all the lineages (*H. vittatus*, *H. tanzaniae* Group D and Group C), except Group B, have high nucleotide and haplotype diversities. This indicates historically large, stable populations or a mix of historically split populations. The low nucleotide and haplotype diversity for Group B could represent an artefact of sampling as this group is represented by two individuals. These findings are partly in contrast with the finding from the *cyt b* data (see section 2.3.3) which indicates a historically stable and large population for *H. vittatus,* Group A and Group B, rapid population growth from a small effective population size for Group C and a severe bottleneck in the recent history of Group D. However, due to the differences in the time window of the two genetic markers, the results represent two points in time of *Hydrocynus* evolution.

Genetic distances between the lineages for both *cyt b* and CR (see section 2.3.3 and 3.3.3) indicate divergence events broadly congruent with those of the phylogenetic trees. The genetic distances within lineages support internal structuring for *H. forskahlii*, *H. tanzaniae* and *H. vittatus* for the Congo and Kwango population. This is congruent with the discoveries of two lineages within *H. forskahlii* in the phylogenetic trees of the *cyt b* data, two lineages within *H. tanzaniae* for the *cyt b* and CR data, and the high haplotype and nucleotide diversity for both Congo and Kwango populations.

4.2.4 Population Structure and Demographic History

All the genealogical sorting index values for each lineage recovered in the phylogenetic trees of both *cyt b* and CR data were 1 ($p<0.005$) (see section 2.3.4 & 3.3.4), which indicates that all the mtDNA lineages recovered in the phylogeny are monophyletic and on independent evolutionary trajectories. The gsi analysis of separate and grouped *H. vittatus* populations indicated a certain degree of mixed history between them. This is consistent with the relatively recent connections between these populations by way of river connections, some of which occur even today (e.g. Okavango and Upper Zambezi) and relatively recent divergence times between several of these populations, which suggests incomplete sorting of genetic variation. This combination of recent and shared history accounts for the poor bootstrap and posterior probability support for several internal branches within the *H. vittatus* lineages on the phylogenetic trees of both data sets (*cyt b* and CR).

A strong degree of geographical structuring of tigerfish into monophyletic mtDNA lineages is shown in several other analyses. The AMOVA (see section 2.3.4 & 3.3.4) showed that significantly more variation is distributed among *Hydrocynus* lineages and populations of *H. vittatus* than is shared within those lineages and populations. This result supports the substantial genetic differentiation recovered in the phylogenies (see section 4.2.2), through the high and well supported gsi values and the high genetic distances (see above & section 4.2.3).

Five of the six populations of *H. vittatus* exhibit evidence of significant demographic changes using tests of selective neutrality (see Section 3.3.4). Explorations of their demographic histories using mismatch distributions gave mixed results. This is probably because the samples represent a large geographical area and exhibit a degree of independent demographic history. This interpretation is supported by the gsi analyses (see above) and the structuring of the networks (see section 4.2.2). Denser sampling within these drainage basins should improve the reliability of the reconstruction of the population demographic histories.

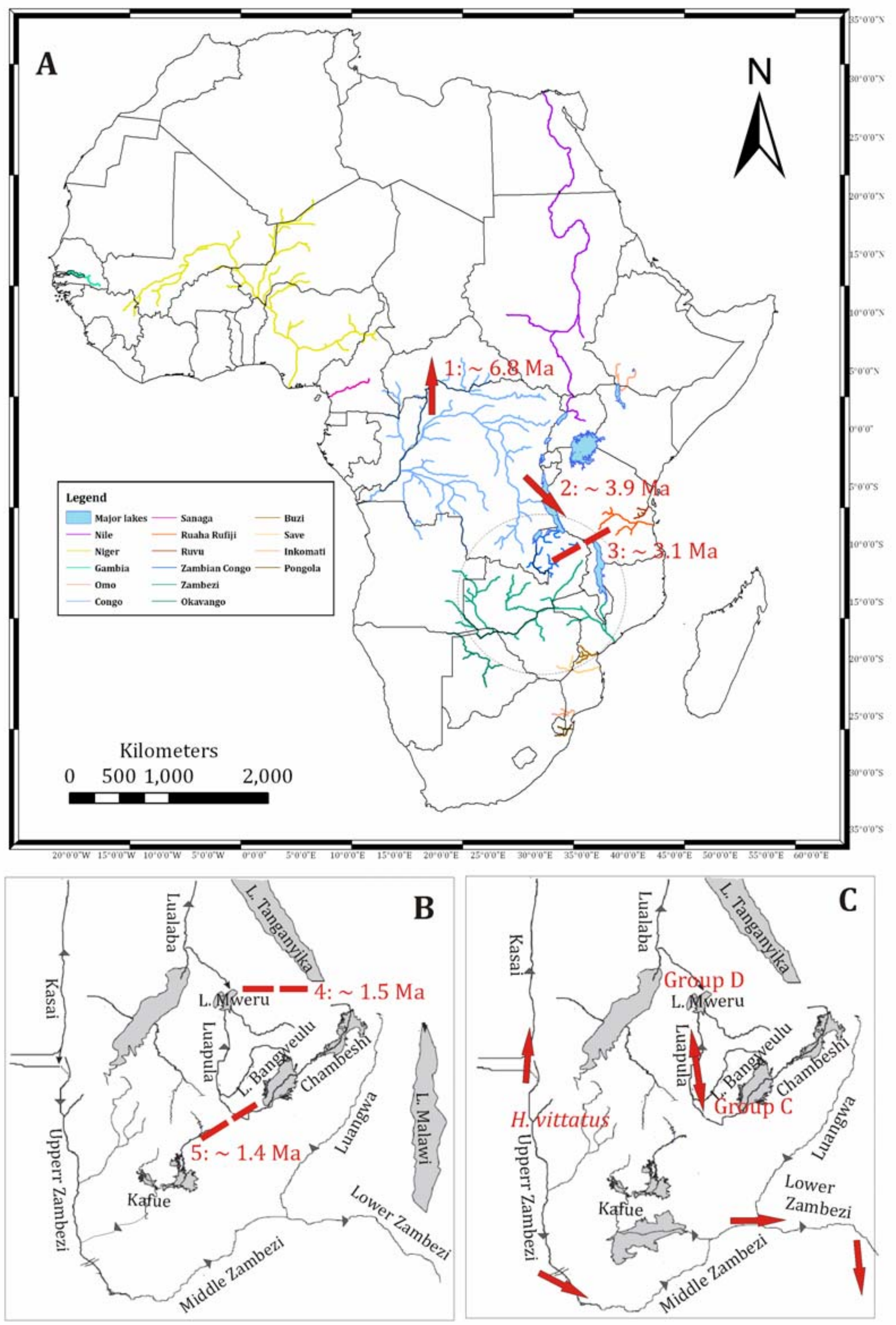
4.3 *Hydrocynus* Evolution and Drainage Reconstruction

The study of the evolution of *Hydrocynus*, discussed in Chapters 2 and 3 and in section 4.2 above, has generated data which can usefully inform reconstructions of drainage evolution. Estimated dates of divergences between lineages constrain the timings of the events that modified drainage topology.

4.3.1 Main Evolutionary Events within *Hydrocynus*

Figures 4.1 and 4.2 present an overview of the biogeographical findings of this study, based on the dated *cyt b* tree, in the context of late Cenozoic drainage evolution. As discussed in Chapter 2 (see section 2.3.2), the confidence intervals around the date estimates are large but these dates can still be used in a broad discussion on the evolutionary history of *Hydrocynus*. *H. goliath* is shown to be the first extant lineage to evolve in the genus *Hydrocynus*. As *H. goliath* occurs only in the Congo system, this implies that this is where *Hydrocynus* first evolved at least 11 Ma.

Dispersal of *Hydrocynus* into north Africa resulted in the founding of *H. brevis*/*H. forskahlii* clade which diverged from a sister Congo lineage approximately 6.8 Ma (No. 1, Figure 4.1a). Such cladogenic events, such as this one between *Hydrocynus* lineages in central and north Africa, highlight former river connections which can be the result of geological events (possibly river captures). These geological events do not appear to have been discussed in the geological literature. The sister relationship between *H. brevis* and *H. forskahlii* recovered in the *cyt b* tree is poorly supported and is not recovered in the other three phylogenetic trees produced which brings this relationship into question (see Section 2.3.2). However, the relationships between individuals of these two species are consistent and well supported in all the phylogenetic trees. *H. forskahlii* can now be called a species complex because genetic data reveal that it diverged into two monophyletic groups about 2.1 Ma which are represented in west African (Sanaga River) and the Nile basins of the Nilo-Sudan ichthyofaunal province. Since no samples were obtained from the west African systems, notably the Niger and Senegal Rivers, the overall ranges of these Nile and Sanaga clades of *H. forskahlii* cannot be resolved in the current study. Within the *H.*



<u>Figure 4.1: Maps indicating main events in the evolutionary history of (a) *Hydrocynus* excluding the Vittatus complex, (b) the Vittatus complex (Groups B, C, D and *H. vittatus*) in the early Pleistocene, and (c) the Vittatus complex in the middle and late Pleistocene to account for their present day distribution</u>

Map constructed using ArcMap 9.3 from the HydroSHEDS digital river database (Lehner *et al.*, 2008) and high-resolution elevation data from the Space Shuttle flight for NASA's SRTM (Jarvis *et al.*, 2006). Numbers indicate events as discussed in the text. Dotted lines indicate reconstructed vicariance/dispersal involving *Hydrocynus* in adjacent drainage basins. Dates indicate the approximate time such vicariance/dispersal occurred calculated using BEAST.

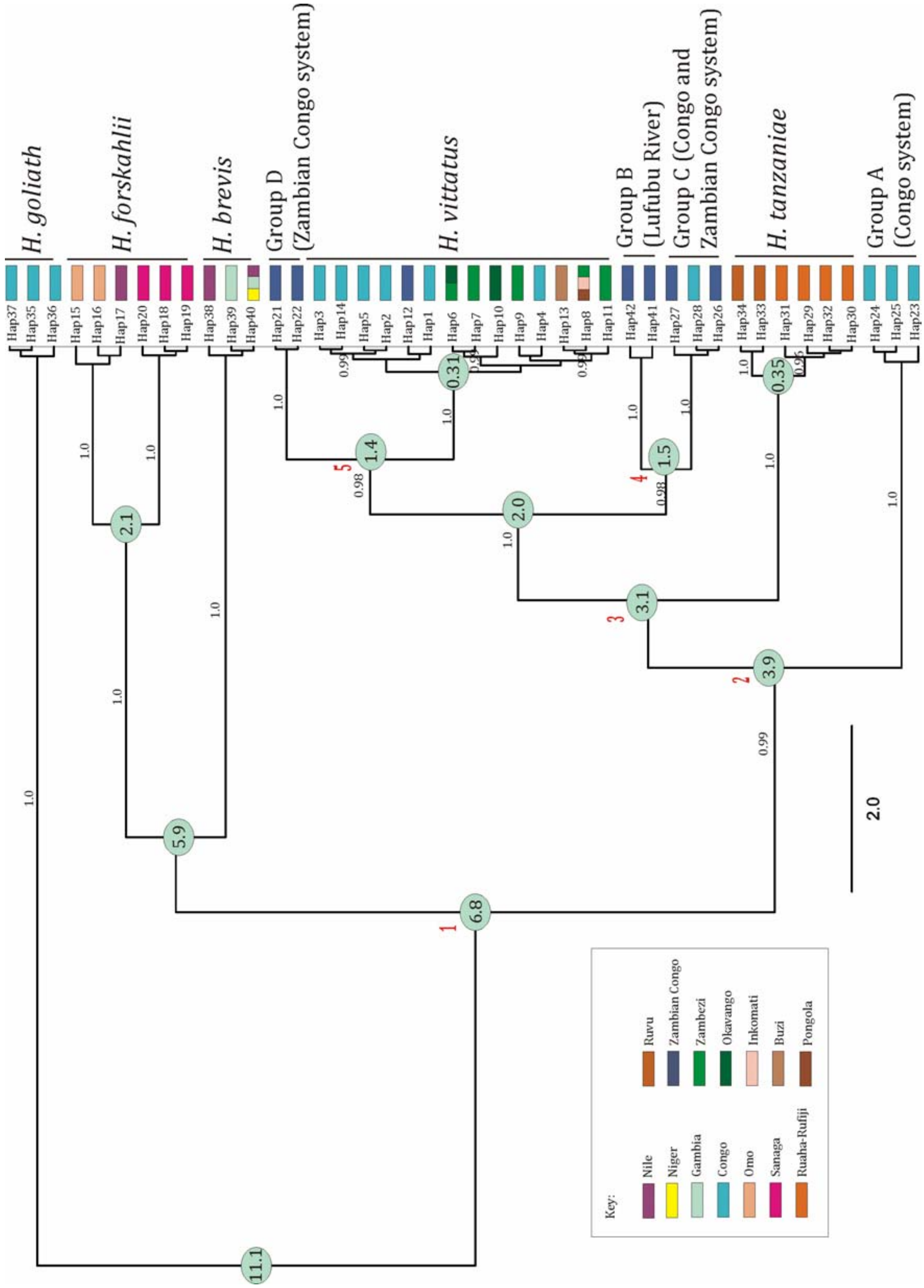


Figure 4.2: Dated Bayesian tree of the *cyt b* sequence data of *Hydrocynus* produced in BEAST
Red numbers above circles indicate events discussed in the text and shown in Figure 4.1. Dates (shown in light green circles) indicate the approximate time such vicariance/dispersal occurred, calculated using BEAST. Only posterior probabilities over 0.70 are shown (small black numbers). The colour key indicates the drainage system(s) represented in each haplotype.

brevis lineage, samples from the Nile, Niger and the Gambia show no divergence, which indicates a recent colonisation event (see Chapter 2, Section 2.3.1).

In the late Neogene, a Congo ancestor sister to the *H. brevis/H. forskahlii* clade gave rise to Group A and the ancestor of *H. tanzaniae* and the Vittatus complex approximately 3.9 Ma (No. 2, Figure 4.1a). The *H. tanzaniae* lineage diverged from the common ancestor of the Vittatus complex about 3.1 Ma (No. 3, Figure 4.1a), possibly as a result of vicariance or dispersal across the Rift Valley in the vicinity of modern Lake Tanganyika. The age of this divergence falls between the time frame of the severing of the connection between the Ruaha-Rufiji and the east-flowing Palaeo-Congo drainage system (40 – 30 Ma) proposed by Stankiewicz and de Wit (2006) and the recolonisation of this drainage area in the past 12 000 years proposed by Roberts (1975) as discussed in Chapter 1 (see Section 1.2.3). The *H. tanzaniae* lineage subsequently diverged between the Rufiji and the Ruvu Rivers approximately 354 ka, making this lineage another possible species complex. Overall, east African rivers contain a high number of endemics, including *H. tanzaniae* (Skelton, 1994). The late Neogene origin of *H. tanzaniae* indicates that the Ruaha-Rufiji basin has possibly acted as a refugium during the dry glacial periods.

A common ancestor of the Vittatus complex lineages (Group B, C and D and *H. vittatus*) evolved approximately 3.1 Ma (Pliocene). The phylogenetic relationships amongst extant *Hydrocynus* lineages indicate that the Vittatus complex was probably seeded from east Africa before the deepening of the Rift Valley, presumably from an ancestor of *H. tanzaniae*. This divergence across the Rift Valley is supported by biogeographical data for other taxa, such as several swampy birds who exhibit a sister relationship between species on either side of the Rift (Cotterill, 2003). The findings of this study suggest that a vicariance or dispersal event isolated tigerfish in the Zambian Congo system approximately 2.0 Ma where they evolved into several lineages. Group C is composed of samples from both Lakes Mweru and Bangweulu and the Dja River in the Congo system. It is likely that this lineage also occurs in the Luapula River, which connects Lakes Mweru and Bangweulu, but only two samples were obtained from this river, both of which represent Group D haplotypes. The major sampling gaps with limited Congo representation make it impossible to resolve the question of where this lineage actually originated, although it may be presumed that it was either in the Congo or the Zambian Congo systems. However, the localized occurrence of Group B (sister clade to Group C) in the Lufubu River (see below), points to a hypothesized origin of Group C in Lake Mweru. Comprehensive sampling across the Congo basin is needed in order to test this interpretation. Moreover, due to the very low proportion of Group C sampled from Lake Bangweulu, it is likely that this lineage moved into Lake Bangweulu relatively recently, presumably from Lake Mweru up the falls on the Luapula River. Such movement is possible in the rainy season (Bell-Cross, 1965a; see Section 1.3). The ancestral sister relationship between the Mweru and Bangweulu members of this lineage and the Dja River sample indicates a past connection across the Congo basin, although further sampling is required to explore this. The sister relationship between Group C and Group B indicates a past connection between the Lufubu and Lake Mweru and the Congo, approximately 1.5 Ma (No. 4, Figure 4.1b). Katongo *et al*. (2005) support a connection between the Lufubu and the Lake Mweru region in the case of the cichlid *Pseudocrenilabrus,* as its sister genus is isolated in the Lufubu River (Koblmüller *et al*., 2008). The geographical proximity of the Lufubu and Lake Mweru and the relative age of the divergence between Groups B and C indicates that Group C probably evolved in Lake Mweru and dispersed into the Congo basin. It is suggested that divergence of Group B followed on isolation of a more widely distributed ancestor in the Proto-Lake Mweru system which became isolated in the Lufubu. This hypothesis can be tested by thorough sampling of both Congo and Zambian Congo populations of tigerfish.

The sister relationship between Group D and *H. vittatus* indicates a split, approximately 1.4 Ma (No. 5, Figure 4.1b), between the Lake Mweru region and the south-central African population today represented in *H. vittatus s.s*. Group D is composed of individuals from Lakes Mweru and Bangweulu, the Luapula River and the Chambeshi River. Due to the proposed origin of the Group C lineage in Lake Mweru at the same time as Group D, it is inferred that this lineage originated in Lake Bangweulu and subsequently spread into Lake Mweru, downstream along the Luapula. Given the genetic data, it seems unlikely that these lakes were completely desiccated during the interpluvials in the Pleistocene or, if they were, there were other refugia for these fish to inhabit. In a phylogeographic study of *Pseudocrenilabrus philander*, a cichlid fish, Katongo *et al*. (2005) recovered separate clades in Lake Mweru and Lake Bangweulu, which supports the findings of the current study of separate *Hydrocynus* lineages in these lakes. Different clades of lechwe antelope are also found around these two lakes, with *Kobus leche* occurring in the Lake Mweru region and *Kobus smithemani* around Lake Bangweulu (Cotterill, 2003 & 2004).

4.3.2 Origin of *H. vittatus*

The phylogenetic and phylogeographic data of the Vittatus complex reveals interesting historical links between drainage systems, which will now be discussed with respect to the origin of *H. vittatus*.

The sister relationship between Group D and *H. vittatus* implies a historical link between the Zambian Congo region and the distribution of *H. vittatus* in south-central Africa, as revealed by the high degree of phylogeographic structuring by river system found within *H. vittatus*. The Congo-Kwango, the Congo-Lufubu and the coastal populations each form groupings and the Okavango and Upper Zambezi show a linked history. The Middle-Lower Zambezi-Shire population group is shown in the network (see Figure 3.7) to be ancestral to the coastal populations to the south.

The findings from the Migrate analyses further inform the emerging drainage reconstruction derived from the evolution of *H. vittatus* (see section 3.3.5). This reconstruction identifies the Upper Zambezi population as the main source population of *H. vittatus* in the analysed drainage basins, suggesting that a vicariant event isolated *Hydrocynus* in the Upper Zambezi, with subsequent post-speciation dispersals into the Congo, the Okavango and the Middle and Lower Zambezi and beyond into the coastal systems. However, it should be noted that unsampled populations can have an effect on the Migrate results (Beerli, 2004) and, due to time and resource constraints on this study, all populations of *H. vittatus* in the study area could not be sampled. Therefore, these results need to be interpreted with caution.

The Migrate findings are unexpected, because they imply movement of *Hydrocynus* via the Palaeo-Chambeshi and Kafue into the Upper Zambezi, resulting in isolation and speciation of *H. vittatus* approximately 300 ka. This is a novel hypothesis which refutes that of Bell-Cross (1965c), who suggests the absence of *H. vittatus* in Upper Kafue basin indicates that this species invaded the Zambezi only after the break-up of the Palaeo-Chambeshi system. The absence of *H. vittatus* in the Upper Kafue could indicate that this species only invaded the Upper Zambezi after the Palaeo-Chambeshi broke up or that it has died out in this river system. Combined evidence, from all analyses conducted above, indicates that the most likely route was from the Zambian Congo region into the Upper Zambezi, via the Kafue, where *H. vittatus* subsequently died out. This extirpation could have been the result of changes in the local climate (see below). This invasion of the Upper Zambezi through the Kafue is also seen in a *P. philander* sub species (Katongo *et al.*, 2005) and in the species flock of serranochromine cichlids (Katongo *et al.*, 2007). *H. vittatus* samples from the whole Zambian Congo region would help to clarify this hypothesis. However, despite a large number of samples being taken from Lakes Mweru and Bangweulu, no *H. vittatus* was sampled. It is possible that *H. vittatus* does not occur in the Zambian Congo, having originated from the Upper Zambezi and moving through the Congo system to populate the Lufubu River. This northward movement of *H. vittatus* into the Congo is in contrast to the hypothesised southward movement from the Congo, via the Kasai, into the Upper Zambezi (Bell-Cross, 1965a; Bowmaker *et al.*, 1978) (see section 1.3.2). The Migrate findings also imply that the Victoria Falls is not a barrier to downstream distribution of *H. vittatus* (see section 1.3.3), at least not in the relatively recent evolutionary past.

Zambian Congo faunal elements are found in the Middle and Lower Zambezi fish fauna, which indicates a past connection between the two systems (Skelton, 1994). This possible connection, in combination with the downstream dispersal of *H. vittatus* over the Victoria Falls from the Upper Zambezi, could have resulted in a population with a mixed history below these falls. Due to the small number of samples and the large number of unsampled populations from the Zambian Congo and the Congo systems, analysis to determine the effects of gene flow on the history of the whole data set of *H. vittatus* could not be conducted. Additional sampling of the Zambian Congo and Congo Rivers is required to conduct meaningful analyses, the results of which could elucidate the complex history of *H. vittatus*.

The variability of past climates, particularly through the Pleistocene (Hewitt, 2004), probably acted on African drainage systems and the links between them. As the diversification in the Vittatus complex occurred during the Plio-Pleistocene, a climatic driver could potentially be the cause of this speciation. Plotting the dates of divergences between lineages onto a high resolution climate curve (e.g. Figure 4 in Lisiecki & Raymo (2005)) could help to answer the question as to whether climate was a determining factor in *Hydrocynus* speciation. However, the confidence intervals of the divergence dates presented in this study are too broad (see section 2.3.2) for a plot to reveal any pattern, as the confidence limits span several oscillations of the climate. Very precise dates of divergence would be required to infer whether climatic factors caused speciation. However, *Hydrocynus* requires and actively seeks out high oxygen conditions and flowing channels (Jackson, 1961; Skelton, 2001). Therefore, these fish would be particularly sensitive to the periods of arid climatic conditions. In addition to tectonic activity that changed drainage topology, the desiccation of rivers and lakes, which severed connections, can help to explain of the vicariant evolution of tigerfish lineages. This arid period could have caused the extirpation of *Hydrocynus* in the Proto-Kafue. The distribution patterns of a species can be affected by its ability to disperse and its life history (Burridge *et al.*, 2008). Such differences are probably the reason for the limited congruence between the evolution of *Hydrocynus* and that of *Synodontis* catfish across tropical Africa (Day *et al.*, 2009).

As it is circular to use fish distributions to infer drainage history only to use the resulting hypothesis of drainage change to explain fish distributions (Bishop, 1995), the hypothesis given here should be applied with caution. It is specific to the history of *H. vittatus* based on the data analysed and interpreted in the framework of published drainage evolution, which is still incomplete.

4.4 Future work

Since this study examined mtDNA, the lineages discussed represent maternal lines of descent. Combined analyses, utilizing nuclear sequence data, in combination with morphological characters, will test whether or not the different mtDNA lineages identified in this study represent novel species in the genus *Hydrocynus*. There is a need for a Pan-African taxonomic revision of *Hydrocynus*, incorporating morphological evidence alongside further genetic analyses, prioritising unsampled populations. Future ecological and behavioural studies of this genus would also help to clarify whether any of the identified lineages occupy different niches, particularly where *Hydrocynus* species occur in sympatry.

While this study is a significant advance in research on *Hydrocynus*, the main sample component (131 samples) was limited to *H. vittatus* in south-central Africa. Even with 131 *H. vittatus* samples, there were gaps in the sampling (e.g. Congo system) and some populations were represented by small sample sizes (e.g. Kwango population). Both of these factors undermined the robustness of some analyses. Due to the almost continent-wide distribution of *Hydrocynus* and the relatively small number of samples from certain populations, considerable work still remains to be done. Further sampling would help to explore the relationships between and within *Hydrocynus* species in greater detail. Additional phylogeographic studies focussing on the species complexes (*H. forskahlii* and *H. tanzaniae*) could reveal interesting relationships between populations of those species. It is very possible that more evolutionarily significant lineages exist than have been identified in this study.

Relatively little of the research conducted on *Hydrocynus* has explicitly addressed the conservation of these fish. Very little is known about population numbers and the conservation status of these populations. Additional work needs to clarify the taxonomic status of the unique evolutionary lineages revealed in this study. Freshwater environments and their fauna are greatly affected by dams, pollution and overfishing (Revenga *et al.*, 2005) and these factors affect the systems inhabited by *Hydrocynus*. For example, the Rufiji River has become a source of hydroelectric power and two dams exist on the Ruaha River (Temple and Sundborg, 1972). Therefore, research into the current status of *Hydrocynus* should be conducted as a matter of urgency in order to assess the impacts on this genus. Since *Hydrocynus* is fished for sport (Skelton, 2001), an improved knowledge of the phylogeography of *H. vittatus*, which resolves the genetic structure of impacted populations, could lead to scientifically-informed regulations aimed at protecting unique lineages and help to decrease over-fishing of certain genotypes, preserving diversity (Charrier *et al.*, 2006). The findings of this study highlight a need to investigate the conservation status of the unique lineages in the Zambian Congo system (especially Lakes Bangweulu and Mweru) as well as in the Congo.

Once a molecular clock has been more precisely calibrated, genetic data can potentially be applied to constrain events in landscape evolution. At the present time, the only calibration point for *Hydrocynus* reflects a minimum age of appearance of the genus. So a generalised mutation rate of *cyt b*, derived from known species of teleosts, was applied in this study and as a result the time estimates of the dated tree are broadly constrained. Obtaining calibration points that could be used to constrain the *Hydrocynus* molecular clock is an important goal for future research on this subject.

4.5 Conclusion

In conclusion, the analyses of the mtDNA *cyt b* and CR sequence data of *Hydrocynus,* examined within a phylogenetic and phylogeographic framework in this study, have described previously unknown patterns of genetic variation in this genus. The hypothesis of drainage evolution in south-central Africa that best explains the current distribution of *H. vittatus* proposed here needs to be tested further using more samples from a wider variety of sample sites across Africa. Further analysis needs to be conducted to expand on these findings. Tighter dating constraints on the *Hydrocynus* phylogeny will improve the precision with which the timings of evolutionary events can be reconstructed.

The phylogeographic knowledge of *Hydrocynus* presented here provides insights into the drainage evolution across tropical Africa and confirms the value of these freshwater fish as biotic indicators of landscape evolution. It is hoped that this work on the evolutionary history of *Hydrocynus* will be extended in the future to contribute further to explaining drainage evolution in south-central Africa and Africa as a whole.

References

Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., Coad, B., Mandrak, N., Balderas, S., Bussing, W., Stiassny, M., Skelton, P., Allen, J., Unmack, P., Naseka, A., Ng, R., Sindorf, N., Robertson, J., Armijo, E., Higgins, J., Heibel, T., Wikramanayake, E., Olson, D., López, H., Reis, R., Lundberg, J., Pérez, M. & Petry, P. 2008. Freshwater ecoregions of the world: A new map of biogeographic units for freshwater biodiversity conservation. *Bioscience*, 58(5):403-414.

Albaret, J., Simier, M., Darboe, F. S., Ecoutin, J., Raffray, J. & Tito de Morais, L. 2004. Fish diversity and distribution in the Gambia Estuary, West Africa, in relation to environmental variables. *Aquatic Living Resources*, 17:35–46.

Allanson, B. R., Hart R. C., O'Keefe J. H. & Roberts R. D. 1990. Inland waters of Southern Africa: An ecological perspective. *Monographiae Biologicae* Vol 64. Kluwer Academic Publishers: Dordrecht.

Avise, J. C. 1995. Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology*, 9(3):686-690.

Avise, J. C. 2000. *Phylogeography: the history and formation of species*. Cambridge: Harvard University Press.

Avise, J. C. 2004. *Molecular markers, natural history and evolution*. Sunderland: Sinauer Associates.

Avise, J. C. 2009. Phylogeography: retrospect and prospect. *Journal of Biogeography*, 36:3–15.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. 1987. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology & Systematics*, 18:489-522.

Ayala, F. J. 1997. Vagaries of the molecular clock. *Proceedings of the National Academy of Sciences of the USA*, 94:7776–7783.

Balon, E. K. 1974. Fishes from the edge of Victoria Falls, Africa: Demise of a physical barrier for downstream invasions. *Copeia*, 1974(3):643-660.

Bandelt, H., Forster, P. & Rohl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular & Biological Evolution*, 16(1):37-48.

Banister, K. E. & Bailey, R. G. 1979. Fishes collected by the Zaire River expedition, 1974-75. *Zoological Journal of the Linnean Society*, 66:205-249.

Baric, S., Salzburger, W. & Sturmbauer, C. 2003. Phylogeography and evolution of the Tanganyikan cichlid genus *Tropheus* based upon mitochondrial DNA sequences. *Journal of Molecular Evolution*, 56:54–68.

Bateman, M. D., Thomas, D. S. G. & Singhvi, A. K. 2003. Extending the aridity record of the Southwest Kalahari: Current problems and future perspectives. *Quaternary International*, 111:37–49.

Bazinet, A. L., Neel, M. C., Shaw, K. L. and Cummings, M. P. In preparation. The genealogical sorting index: Software and web site for quantifying exclusivity of lineages. Available at http://gsi.umiacs.umd.edu/documentation/

Beerli, P. 2004. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology,* 13:827-836.

Beerli, P. 2008. Migrate version 3.0 - a maximum likelihood and Bayesian estimator of gene flow using the coalescent. Available at http://popgen.scs.edu/migrate.html.

Beerli, P. & Felsenstein, J. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, 152(2):763-73.

Beheregaray, L. B. 2008. Twenty years of phylogeography: The state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, 17(17):3754-3774.

Bell-Cross, G. 1965a. The distribution of fishes in central Africa. *Fisheries Research Bulletin (Zambia)*, 4:3-20.

Bell-Cross, G. 1965b. Preliminary observations on *Hydrocynus vittatus* in the Upper Zambezi river system. *Fisheries Research Bulletin (Zambia)*, 4:21-26.

Bell-Cross, G. 1965c. Physical barriers separating the fishes of the Kafue and Middle Zambezi River Systems. *Fisheries Research Bulletin (Zambia)*, 4:119.

Bell-Cross, G. 1972. The fish fauna of the Zambezi system. *Arnodia (Rhodesia)*, 5(29):1–19.

Bell-Cross, G. 1973. The fish fauna of the Buzi River system in Rhodesia and Mozambique. *Arnodia (Rhodesia),* 6(5): 1-14.

Bensasson, D., Zhang, D., Hartl, D. L. & Hewitt, G. M. 2001. Mitochondrial pseudogenes: Evolution’s misplaced witnesses. *Trends in Ecology & Evolution*, 16(6):314-321.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. 2005. Genbank. *Nucleic Acids Research*, 33:D34–D38.

Berendzen, P. B., Gamble T. & Simons, A. M. 2008. Phylogeography of the bigeye chub *Hybopsis amblops* (Teleostei: Cypriniformes): Early Pleistocene diversification and post-glacial range expansion. *Journal of Fish Biology*, 73:2021–2039.

Bermingham, E. & Avise, J. C. 1986. Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics*, 113:939-965.

Billington, N. & Herbert, P. D. N. 1991. Mitochondrial DNA diversity in fishes and its implications for introductions. *Canadian Journal of Fisheries & Aquatic Sciences*, 48(Supplementary 1):80–94.

Bishop, P. 1995. Drainage rearrangement by river capture, beheading and diversion. *Progress in Physical Geography*, 19(4):449-473.

Boelhouwers, J.C. & Meiklejohn, K. I. 2002. Quaternary periglacial and glacial geomorphology of southern Africa: Review and synthesis. *South African Journal of Science*, 98:47-55.

Bowen, B. W. & Grant, W. S. 1997. Phylogeography of the sardines (*Sardinops* spp.): Assessing biogeographic models and population histories in temperate upwelling zones. *Evolution*, 51(5):1601-1610.

Bowie, R. C. K., Fjeldså, J., Hackett, S. J., Bates, J. M. & Crowe T. M. 2006. Coalescent models reveal the relative roles of ancestral polymorphism, vicariance, and dispersal in shaping phylogeographical structure of an African montane forest robin. *Molecular Phylogenetics & Evolution*, 38:171–188.

Bowmaker, A. P. M. 1973. *An hydrobiological study of the Mwenda River and its mouth, Lake Kariba*. PhD Thesis, University of the Witwatersrand, Johannesburg.

Bowmaker, A. P., Jackson, P. N. & Jubb, R. A. 1978. Freshwater fishes. *In* Werger, M. J. A. (ed.), *Biogeography & Ecology of Southern Africa*. The Hague: Dr W. H. Junk, p. 1181–1230.

Brewster, B. 1986. A review of the genus *Hydrocyuns* Cuvier 1819 (Teleostei: Characiformes). *Bulletin of the British Museum (Natural History), Zoology*, 50(3):163-206.

Bromham, L. & Penny, D. 2003. The modern molecular clock. *Nature*, 4:216-224.

Brown, W. M., George, M. & Wilson, A. C. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the USA,* 76(4):1967-1971.

Bryant, D. & Moulton, V. 2004. NeighborNet: An agglomerative algorithm for the construction of planar phylogenetic networks. *Molecular & Biological Evolution*, 21:255–265.

Burke, K. 1996. The African plate. *South African Journal of Geology*, 99:339-410.

Burke, K. & Gunnel, Y. 2008. *The African erosion surface: a continental-scale synthesis of geomorphology*. Colorado: Geological Society of America.

Burridge, C. P., Craw, D., Jack, D. C., King, T. M. & Waters, J. M. 2008. Does fish ecology predict dispersal across a river drainage divide? *Evolution*, 62(6):1484–1499.

Burridge, C. P., Craw, D. & Waters, J. M. 2007. An empirical test of freshwater vicariance via river capture. *Molecular Ecology*, 16:1883-1895.

Burridge, C. P., Meléndez, R., & Dyer, B. S. 2006. Independent origins of the Juan Fernández kelpfish fauna (Perciformes: Chironemidae), and evidence for frequent and unidirectional dispersal of cirrhitoid fishes across the South Pacific. *Systematic Biology,* 55:566-578.

Calcagnotto, D., Schaefer, S. A. & DeSalle, R. 2005. Relationships among characiform fishes inferred from analysis of nuclear and mitochondrial gene sequences. *Molecular Phylogenetics & Evolution*, 36:135–153.

Charrier, G., Chenel, T., Durand, J. D., Girand, M., Quiuiou, L. & Laroche, J. 2006. Discrepancies in phylogeographical patterns of two European anglerfishes (*Lophius budegassa* and *Lophius piscatorius*). *Molecular Phylogenetics and Evolution*, 38:742–754.

Chow, S. & Hazama, K. 1998. Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology*, 7:1247-1263.

Cooke, G. M., Chao, N. L. & Beheregaray, L. B. 2009. Phylogeography of a flooded forest specialist fish from central Amazonia based on intron DNA: The cardinal tetra *Paracheirodon axelrodi*. *Freshwater Biology,* 54(6):1216-1232.

Cotterill, F. P. D. 2003. Geomorphological influences on vicariant evolution in some African mammals in the Zambezi basin: Some lessons for conservation. *In* Plowman, A. (ed.), *Ecology and Conservation of Small Antelope*. Fürth: Filander Verlag, p. 11-58.

Cotterill, F. P. D. 2004. Drainage evolution in south-central Africa and vicariant speciation in swamp-dwelling weaver birds and swamp flycatchers. *The Honeyguide*, 25(1):7-25.

Cotterill, F. P. D. 2006. *The evolutionary history and taxonomy of the Kobus leche species complex of south-central Africa in the context of palaeo-drainage dynamics*. PhD Thesis, University of Stellenbosch, Stellenbosch.

Cotterill, F. P. D. & Goodier, S. 2008. DNA clocks for dating landforms. *Quest*, 4(2):19-21.

Craw, D., Burridge, C., Norris, R. & Waters, J. 2008. Genetic ages for Quaternary topographic evolution: A new dating tool. *Geology,* 36(1):19-22.

Cummings, M. P., Neel, M. C. & Shaw, K. L. 2008. A genealogical approach to quantifying lineage divergence. *Evolution,* 62(9):2411-2422.

Daget, J. 1960. La faune ichtyologique du bassin de la Gambie. *Bulletin de l'Institut Français d'Afrique Noire (A),* 22(2):610-619.

Day, J. J., Bills, R. & Friel, J. P. 2009. Lacustrine radiations in African Synodontis catfish. *Journal of Evolutionary Biology,* 22:805-817.

Debenham, F. 1947. The Bangweulu swamps of Central Africa. *Geographical Review,* 37(3):351-368.

de Wit, M. 2007. The Kalahari epeirogeny and climate change: Differentiating cause and effect from core to space. *South African Journal of Geology,* 110(2-3):367-392.

Derricourt, R. M. 1976. Retrogression rate of the Victoria Falls and the Batoka Gorge. *Nature,* 264(5581):23-25.

Dixey, F. 1943. The morphology of the Congo-Zambesi Watershed. *South African Geographical Journal,* 25:20-41.

Dixey, F. 1944. African landscape. *Geographical Review,* 34(3):457-465.

Dixey, F. 1945. The geomorphology of Northern Rhodesia. *Transactions of the Geological Society of South Africa,* 48:9-45.

Dollar, E. S. J. 1998. Palaeofuvial geomorphology in southern Africa: A review. *Progress in Physical Geography,* 22(3):325-349.

Domingues, V. S., Santos, R. S., Brito, A., Alexandrou, M. & Almada, V. C. 2007. Mitochondrial and nuclear markers reveal isolation by distance and effects of Pleistocene glaciations in the northeastern Atlantic and Mediterranean populations of the white seabream (*Diplodus sargus*, L.). *Journal of Experimental Marine Biology and Ecology,* 346(1-2):102-113.

Doucouré, C. M. & de Wit, M. J. 2003. Old inherited origin for the present near-bimodal topography of Africa. *Journal of Earth Science,* 36:371-388.

Doyle, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. *Systematic Biology,* 17(1):144-163.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology,* 4(5):699-710.

Drummond, A. & Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology,* 7:214.

du Toit, A. L. 1927. The Kalahari. *South African Journal of Science*, 24:88-101.

du Toit, A. L. 1933. Crustal movement as a factor in the geographical evolution of South Africa. *South African Geographical Journal*, 16:3-20.

Excoffier, L., Laval, G. & Schneider, S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1:47–50.

Excoffier, L., Smouse, P. E. & Quattro, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, 31:479-491.

Farias, I. P., Ortí, G., Sampaio, I., Schneider, H. & Meyer, A. 2001. The cytochrome b gene as a phylogenetic marker: The limits of resolution for analyzing relationships among cichlid fishes. *Journal of Molecular Evolution*, 53(2):89-103.

Fu, Y. X. 1996. New statistical tests of neutrality for DNA samples from a population. *Genetics*, 143:557-570.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147:915–925.

Gabie, V. 1965. Problems associated with the distribution of freshwater fishes in southern Africa. *South African Journal of Science*, 61(11):383-391.

Gagiano, C. L., Steyn, G. J. & du Preez, H. H. 1996. Tooth replacement of tigerfish *Hydrocynus vittatus* from the Kruger National Park. *Koedoe*, 39(1):117-122.

Gaigher, I. G. 1970. Ecology of the tiger-fish (*Hydrocynus vittatus*) in the Incomati River system, South Africa. *Zoologica Africana*, 5(2):211-227.

Gaigher, I. G. & Pott, R. M. C. 1973. Distribution of fishes in southern Africa. *South African Journal of Science*, 69:25-27.

Gasse, F. 2000. Hydrological changes in the African tropics since the Last Glacial Maximum. *Quaternary Science Reviews,* 19:89-211.

Giresse, P. 2005. Mesozoic–Cenozoic history of the Congo Basin. *Journal of African Earth Sciences*, 43:301–315.

Goudie, A. S. 2005. The drainage of Africa since the Cretaceous. *Geomorphology*, 67:437-456.

Grove, A. T. 1969. Landforms and Climatic Change in the Kalahari and Ngamiland. *The Geographical Journal*, 135(2):191-212.

Grubb, P. 1999. Evolutionary processes implicit in distribution patterns of modern African mammals. *In* Bromage, T. G. & Schrenk, F. (eds.), *African Biogeography, Climate Change, and Human Evolution*. Oxford: Oxford University Press, p. 150-166.

Gumbricht, T., McCarthy, T.S. & Merry, C.L. 2001. The topography of the Okavango Delta, Botswana, and its tectonic and sedimentological implications. *South African Journal of Geology*, 104:243–264.

Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41:95-98.

Harpending, H. C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, 66(4):591-600.

Hasegawa, M., Kishino, H. & Yano, T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22:160-174.

Hewitt, G. 2004. The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology*, 1:4. 10.1186/1742–10.1186/9994–1–4.

Hey, J. & Nielsen, R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, 167:747-760.

Hrbek, T., Stölting, K. N., Bardakci, F., Küçük, F., Wildekamp, R. H. & Meyer, A. 2004. Plate tectonics and biogeographical patterns of the *Pseudophoxinus* (Pisces: Cypriniformes) species complex of central Anatolia, Turkey. *Molecular Phylogenetics and Evolution*, 32(1):297-308.

Huelsenbeck, J. P. & Ronquist, F. 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics*, 17:754–755.

Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, 294:2310-2314.

Hurwood, D. A. & Hughes, J. M. 1998. Phylogeography of the freshwater fish, *Mogurnda adspersa*, in streams of northeastern Queensland, Australia: Evidence for altered drainage patterns. *Molecular Ecology,* 7(11):1507-1517.

Huson, D. H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology & Evolution,* 23(2):254-267. Software available at www.splitstree.org.

Irwin, D. M., Kocher, T. D. & Wilson, A. C. 1991. Evolution of the cytochrome b gene of mammals. *Journal of Molecular Evolution*, 32:128-144.

Isaac, G. L. & Behrensmeyer, A. K. 1997. Geological context and palaeoenvironments. *In* Isaac, G. L. (ed.), *Koobi Fora Research Project, Volume 5, Plio-Pleistocene Archaeology.* Oxford: Clarendon Press, p. 12–70.

Ivanova, N. V., Zemlak, T. S., Hanner, R. H. & Herbert, P. D. N. 2007. Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes,* 7:544-548.

Jackson, P. B. N. 1961. The impact of predation, especially by the tiger-fish (*Hydrocyon vittatus* Cast.) on African freshwater fishes. *Proceedings of the Zoological Society of London*, 136:603-622.

Jackson, P. N. B. 1986. Fish of the Zambezi system. *In* Davies, B. R. & Walker, K. F. (eds.), *The Ecology of River Systems*. The Hague: Dr W. H. Junk, p. 269–288.

Jackson, P. B. N. 2000. Freshwater fishery research organizations in central and eastern Africa – a personal recollection. *Transactions of the Royal Society of South Africa*, 55:1–81.

Jarvis A., Reuter, H. I., Nelson, A. & Guevara, E. 2006. Hole-filled SRTM for the globe Version 3. Available from the CGIAR-CSI SRTM 90m Database at http://srtm.csi.cgiar.org.

Johansen, S., Guddal, P. H. & Johansen, T. 1990. Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua*. *Nucleic Acids Research,* 18(3):411-419.

Johns, G. C. & Avise, J. C. 1998. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Molecular & Biological Evolution*, 15(11):1481–1490.

Johnson, D. L. 2002. Darwin would be proud: Bioturbation, dynamic denudation, and the power of theory in science. *Geoarchaeology,* 17(1):7–40.

Jubb, R. A. 1952. A note on the distribution of the *Hydrocyon vittatus,* Castilnau, the well known tigerfish of the Zambezi and Limpopo rivers. *South African Journal of Science*, 49:50–51.

Jubb, R. A. 1977. Comments on Victoria Falls as a physical barrier for downstream dispersal of fishes. *Copeia*, 1:198-199.

Katongo, C., Koblmüller, S., Duftner, N., Makasa, L. & Sturmbauer, C. 2005. Phylogeography and speciation in the *Pseudocrenilabrus philander* species complex in Zambian Rivers. *Hydrobiologia*, 542:221–233.

Katongo, C., Koblmüller, S., Duftner, N., Mumba, L. & Sturmbauer, C. 2007. Evolutionary history and biogeographic affinities of the serranochrome cichlids in Zambian rivers. *Molecular Phylogenetics & Evolution*, 45:326-338.

Kenmuir, D. H. S. 1973. The ecology of the tigerfish, *Hydrocynus vittatus* Castelnau, in Lake Kariba. *Occasional Papers of the National Museums & Monuments of Rhodesia,* B5(3):115-170.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature,* 217:624-626.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution,* 16:111-120.

Koblmüller, S., Schliewn, U. K., Duftner, N. & Sefc, K. M. 2008. Age and spread of the haplochromine cichlid fishes in Africa. *Molecular Phylogenetics & Evolution,* 49:153–169.

Kreiser, B. R., Mitton, J. B. & Woodling, J. D. 2001. Phylogeography of the plains killifish, *Fundulus zebrinus*. *Evolution,* 55(2):339-350.

Lavier, L. L., Steckler, M. S. & Brigaud, F. 2001. Climatic and tectonic control on the Cenozoic evolution of the West African margin. *Marine Geology,* 178(1):63-80.

Lee, W., Conroy, J. Howell, W. H. & Kocher, T. D. 1995. Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution,* 41:54-66.

Lehner, B., Verdin, K. & Jarvis, A. 2008. New global hydrography derived from spaceborne elevation data. *EOS, Transactions, American Geophysical Union,* 89(10):93-94.

Lessa, E. P. 1992. Rapid surveying of DNA sequence variation in natural populations. *Molecular & Biological Evolution,* 9(2):323-330.

Lévêque, C. 1997. *Biodiversity dynamics and conservation: the freshwater fish of tropical Africa*. United Kingdom: Cambridge University Press.

Li, W. H. 1977. Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics,* 85:331-337.

Lisiecki, L. E. & Raymo, M. E. 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records. *Paleoceanography*, 20:PA1003. doi:10.1029/2004PA001071

Lopez, J. V., Yuhki, N., Masuda, R., Modi, W. & O'Brien, S. J. 1994. Numt: A recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution,* 39(2):174-190.

Lyons, R. P., Scholz, C. A., Buoniconti, M. R. & Martin, M. R. 2009. Late Quaternary stratigraphic analysis of the Lake Malawi Rift, East Africa: An integration of drillcore and seismic-reflection data. *Palaeogeography, Palaeoclimatology, Palaeoecology,* In press, corrected proof available at http://www.sciencedirect.com/science/journal/00310182.

Malaisse, F. 1997. *Se nourrir en forêt claire Africaine: approche écologique et nutrionnelle*. Gembloux: Presses Agronomiques de Gembloux.

McClanahan, T. R. & Young, T. P. 1996. *East African ecosystems and their conservation*. New York: Oxford University Press.

Meintjies, M. 1999. *Zambezi Tiger: Flyfishing for Tiger fish on the Upper Zambezi River where four countries meet*. Edenvale: Entrepot Books.

Meintjies, M. 2005. *Remarkable flyfishing destinations of Southern Africa*. Cape Town: Struik.

Modolo, L., Salzburger, W. & Martin, R. D. 2005. Phylogeography of Barbary macaques (*Macaca sylvanus*) and the origin of the Gibraltar colony. *Proceedings of the National Academy of Sciences of the USA,* 102(20):7392–7397.

Moore, A. E. 1999. A reappraisal of epeirogenic flexure axes in Southern Africa. *South African Journal of Geology,* 102(4):363-376.

Moore, A. E., Blenkinsop, T. & Cotterill, F. P. D. 2009a. Southern African topography and erosion history: plumes or plate tectonics? *Terra Nova*, 21:310–315.

Moore, A. E. & Cotterill, F. P. D. 2009. Victoria Falls: Mosi-oa-Tunya - the Smoke that Thunders. *In* Migon, P. (ed.), *Geomorphological Landscapes of the World*. Berlin: Springer, p. 143-153.

Moore, A. E., Cotterill, F. P. D., Broderick, T. & Plowes, D. 2009b. Landscape evolution in Zimbabwe from the Permian to present, with implications for kimberlite prospecting. *South African Journal of Geology*, 112:65-88.

Moore, A. E. & Larkin, P. A. 2001. Drainage evolution in south-central Africa since the breakup of Gondwana. *South African Journal of Geology*, 104:47-68.

Morrison, C. L., Lemarié, D. P., Wood, R. M. & King, T. L. 2006. Phylogeographic analyses suggest multiple lineages of *Crystallaria asprella* (Percidae: Etheostominae). *Conservation Genetics*, 7(1):129-147.

Murphy, W. J. & Collier, G. E. 1997. A molecular phylogeny for Aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): The role of vicariance and the origins of annualism. *Molecular & Biological Evolution,* 14(8):790-799.

Næsje, T. F., Hay, C. J., Kapirika, S., Sandlund, O. T. & Thorstad, E. B. 2001. *Some ecological and socio-ecological impacts of an angling competition in the Zambezi River, Namibia* (*NINA-NIKU Project Report 14)*. Trondheim: Norwegian Institute for Nature Research.

Nichols, R. 2001. Gene trees and species trees are not the same. *Trends in Ecology & Evolution,* 16:358-364.

Nielsen, R. & Wakeley, J. 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics,* 158:885–896.

Ogg, G. 2009. *International stratigraphic chart 2009*. Available at www.stratigraphy.org.

Partridge, T. C. 1997. Cenozoic environmental change in southern Africa, with special emphasis on the last 200,000 years. *Progress in Physical Geography*. 21(1):3–22.

Paugy, D. & Guégan J. F. 1989. Note à propos de trios espèces d'*Hydrocynus* (Pisces, Characidae) du basin du Niger suivie de la rehabilitation de l'espèce *Hydrocynus vittatus* (Castelnau, 1861). *Revue d'Hydrobiologie Tropicale,* 22(1): 63-69.

Peters, C. R. & O'Brien, E. M. 2001. Palaeo-lake Congo: Implications for Africa's late Cenozoic climate – some unanswered questions. *In* Heine, K. & Jürge, R. (eds.), *Proceedings of the VXth INQUA Conference: Durban, South Africa, 3-11 August 1999*. Lisse: Swets & Zeiglinger, p. 11-18.

Pfenninger, M. & Posada, D. 2002. Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylommatophora): Fragmentation, corridor migration, and secondary contact. *Evolution,* 56(9):1776-1788.

Posada, D. 1999. Collapse Version 1.1. Available at http://crandalllab.byu.edu/Computer.aspx.

Posada, D. & Crandall, K. A. 1998. Modeltest: Testing the model of DNA substitution. *Bioinformatics,* 14(9):817-818.

Posada, D. & Crandall, K. A. 2001. Selecting models of nucleotide substitution: An application to Human Immunodeficiency Virus 1 (HIV-1). *Molecular & Biological Evolution,* 18(6):897–906.

Potter, P. E. & Hamblin, W. K. 2006. *Big rivers worldwide*. Provo: Brigham Young University.

Reeves, C. & de Wit, M. 2000. Making ends meet in Gondwana: retracing the transforms of the Indian Ocean and reconnecting continental shear zones. *Terra Nova*, 12(6):272-280.

Reid, J. M. 1996. Ichthyogeography of the Guinea-Congo rain forest, West Africa. *Proceedings of the Royal Society of Edinburgh,* 104(B):285-312.

Revenga, C., Campbell, I., Abell, R. de Villiers, P. & Bryer, M. 2005. Prospects for monitoring freshwater ecosystems towards the 2010 targets. *Philosophical Transactions of the Royal Society B,* 360(1454):397-413.

Richly, E. & Leister, D. 2004. Numts in sequenced eukaryotic genomes. *Molecular & Biological Evolution,* 21(6):1081-1084.

Roberts, T. R. 1975. Geographical distribution of African freshwater fishes. *Zoological Journal of the Linnean Society,* 57:249-319.

Rocha-Olivares, A., Garber, N. M., Garber, A. F. & Stuck, K. C. 2005. Structure of the mitochondrial control region and flanking tRNA genes of *Mugil cephalus*. *Hidrobiológica,* 15(2):139-149.

Rogers, A. R. 1995. Genetic evidence for a Pleistocene population explosion. *Evolution,* 49(4):608-615.

Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular & Biological Evolution,* 9:552-569.

Roux, F. & van Buynder, D. 2003. Inkomati Tigerfish Telemetry Project. *In Kruger National Park Project Reports 2003*. Johannesburg: South African National Parks, p. 110-111.

Runge, J. 2007. The Congo River, Central Africa. (*In* Gupta, A. (ed.), *Large rivers: geomorphology and management*. Wiley & sons, Chichister. p. 293-310.)

Russel, A. L., Ranivo, J., Palkovacs, E. P. Goodman, S. M. & Yoder, A. D. 2007. Working at the interface of phylogenetics and population genetics: A biogeographical analysis of *Triaenops* spp. (Chiroptera: Hipposideridae). *Molecular Ecology,* 16:839–851.

Shahin, M. 2002. *Hydrology and water resources of Africa*. Dordrecht: Kluwer Academic publishers.

Skelton, P. H. 1988. The distribution of African freshwater fishes. *In* Lévêque, C., Bruton M. N. & Ssentongo G. W. (eds.), *Biologie et écologie des poisson d'eau douce africans*. Paris: ORSTOM, p. 65-91.

Skelton, P. H. 1990. Scientific nomenclature of tigerfish – rehabilitation of *Hydrocynus vittatus*. *Ichthos,* 26:7.

Skelton, P. H. 1994. Diversity and distribution of freshwater fishes in East and Southern Africa. *Annales du Musée de la République de l'Afrique Centrale, Sciences Zoologiques,* 275:95-131.

Skelton, P. H. 2001. *A complete guide to the freshwater fishes of Southern Africa*. Cape Town: Struik.

Sokal, R. R. & Sneath, P. H. A. 1963. *Principles of numerical taxonomy*. San Francisco: Freeman.

Sorenson, M. D. & Quinn, T. W. 1998. Numts: A challenge for avian systematics and population biology. *The Auk,* 115(1):214-221.

Stankiewicz, J. & de Wit, M. J. 2006. A proposed drainage evolution model for Central Africa - Did the Congo flow east? *Journal of African Earth Sciences,* 44:75–84.

Stewart, K. M. 2001. The freshwater fish of Neogene Africa (Miocene–Pleistocene): Systematics and biogeography. *Fish & Fisheries,* 2:177–230.

Stiassny, M. L. J. & Schliwen, U. K. 2007. Congochromis, a new cichlid genus (Teleostei: Cichlidae) from central Africa, with the description of a new species from the Upper Congo River, Democratic Republic of Congo. *American Museum Novitates,* no. 3576.

Sturmbauer, C., Baric, S., Salzburger, W., Ruber, L. & Verheyen, E. 2001. Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Molecular & Biological Evolution,* 18(2):144–154.

Swofford, D. L. 1998. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods)*. Version 4. Sinauer Associates, Sunderland, Mass.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics,* 123:585-595.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology & Evolution,* 24:1596-1599.

Tamura, K. & Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular & Biological Evolution,* 10(3):512-526.

Tang, Q., Liu, H., Mayden, R. & Xiong, B. 2006. Comparison of evolutionary rates in the mitochondrial DNA cytochrome b gene and control region and their implications for phylogeny of the Cobitoidea (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution,* 39(2):347-357.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. (*In* Miura, M. M. (ed.), *Lectures on mathematics in the life sciences*. American Mathematical Society, New York. p. 57-86.)

Technelysium, 2006. Chromas 2.13. Technelysium Pty Ltd. Available at www.technelysium.com.au/chromas.html.

Temple, P. H. & Sundborg, Å. 1972. The Rufiji River, Tanzania, hydrology and sediment transport. *Geografiska Annaler, Series A, Physical Geography,* 54(3):345-368.

Thieme, M. L., Abell, R., Stiassny, M. L. J. S., Skelton, P., Burgess, N., Lehner, B., Dinerstein, E., Olson, D., Teugels, G. & Kamdem-Toham, A. 2005. *Freshwater ecoregions of Africa and Madagascar: A conservation assessment*. Washington: Island Press.

Thomas, D. S. G & Shaw, P. A. 1991. *The Kalahari Environment*. Cambridge: Cambridge University Press.

Thomas, D. S. G & Shaw, P. A. 2002. Late Quaternary environmental change in central southern Africa: New data, synthesis, issues and prospects. *Quaternary Science Reviews,* 21:783-797.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22:4673-4680.

Thorstad, E. B., Hay, C. J., Næsje, T. F., Chanda, B. & Økland, F. 2002. *Movements and habitat utilisation of tigerfish (Hydrocynus vittatus) in the Upper Zambezi River: Implications for fisheries management (NINA Project Report 19)*. Trondheim: Norwegian Institute for Nature Research.

Tweddle, D., Lewis, D. S. C. & Willoughby, N. G. 1979. The nature of the barrier separating the Lake Malawi and Zambezi fish faunas. *Ichthyological Bulletin of Rhodes University,* 39:1-9.

van der Bank, F. H., Kotzé, A. & Falk, T. M. 2005. Phylogenetic relationships between some African Characidae, Distichodontidae and Hepsetidae: Results of allozyme and mtDNA data analyses. *African Journal of Aquatic Sciences,* 30(1):77–84.

Walsh, P. S., Metzger, D. A. & Higuchi, R. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques,* 10(4):506-513.

Watchman, A. L. & Twidale, C.R. 2002. Relative and 'absolute' dating of land surfaces. *Earth-Science Reviews,* 58:1–49.

Waters, J. M., Rowe, D. L., Apte, S., King, T. M., Wallis, G. P., Anderson, L., Norris, R. J., Craw, D. & Burridge, C. P. 2007. Geological dates and molecular rates: Rapid divergence of rivers and their biotas. *Systematic Biology,* 56(2):271-282.

Waters, J. M. & Wallis, G. P. 2001. Mitochondrial DNA phylogenetics of the *Galaxias vulgaris* complex from South Island, New Zealand: Rapid radiation of a species flock. *Journal of Fish Biology,* 58:1166–1180.

Wellington, J. H. (1955) *Southern Africa: A Geographical Study, Volume 1: Physical Geography*. Cambridge: Cambridge University Press.

Woldegabriel, G., Heiken, G., White, T. D., Asfaw, B., Hart, W. K. & Renne, P. R. 2000. Volcanism, tectonism, sedimentation, and the palaeoanthropological record in the Ethiopian Rift System. *Geological Society of America Special Papers*, 345:83–99.

Wright, S. 1951. The genetical structure of populations. *Annals of Eugenics,* 15:323-354.

Zhang, D. X. & Hewitt, G. M. 2003. Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects. *Molecular Ecology,* 12(3):563-84.

Zuckerkand, E. & Pauling, L. 1962. Molecular disease, evolution, and genetic heterogeneity. *In* Kasha, M. & Pullman, B. (eds.), *Horizons in Biochemistry.* New York: Academic Press, p. 189-225.

Appendices

Appendix 1: Table of samples, their haplotype designations and their collection sites

Sample	*Cyt b* Haplotype No.	CR Haplotype No.	Species	Country	River/Lake
Brev 1	Hap 38	-	*H. brevis*	Sudan	Nile River
Brev 2	Hap 40	-	*H. brevis*	Sudan	Nile River
Brev 3	Hap 40	-	*H. brevis*	Sudan	Nile River
Brev 5	Hap 40	-	*H. brevis*	Sudan	Nile River
Brev 7	Hap 38	-	*H. brevis*	Sudan	Nile River
Gam 1	Hap 39	-	*H. brevis*	Senegal	Gambia River
Gam 2	Hap 40	-	*H. brevis*	Senegal	Gambia River
Gam 3	Hap 40	-	*H. brevis*	Senegal	Gambia River
AY791400	Hap 40	-	*H. brevis*	Niger	Niger River
Ethio 1	Hap 15	-	*H. forskahlii*	Ethiopia	Lake Chamo
Ethio 2	Hap 16	-	*H. forskahlii*	Ethiopia	Lake Chamo
Ethio 3	Hap 16	-	*H. forskahlii*	Ethiopia	Lake Chamo
Saca 1	Hap 19	-	*H. forskahlii*	Cameroon	Sanaga River
Saca 2	Hap 20	-	*H. forskahlii*	Cameroon	Sanaga River
Saca 3	Hap 19	-	*H. forskahlii*	Cameroon	Sanaga River
Saca 5	Hap 18	-	*H. forskahlii*	Cameroon	Sanaga River
SUD 3	Hap 17	-	*H. forskahlii*	Sudan	White Nile River
CAR 1	Hap 35	-	*H. goliath*	Central African Republic	Congo River
Co 1	Hap 35	-	*H. goliath*	Congo	Congo River
Co 6	Hap 36	-	*H. goliath*	Congo	Congo River
Co 207	Hap 35	-	*H. goliath*	Congo	Congo River
Hv 200	Hap 35	-	*H. goliath*	Congo	Congo River
Hv 205	Hap 35	-	*H. goliath*	Congo	Congo River
INGA 1	Hap 37	-	*H. goliath*	Congo	Congo River
Kw 2a	Hap 36	-	*H. goliath*	Congo	Kwango River
Ruv 1	Hap 33	Hap 116	*H. tanzaniae*	Tanzania	Ruvu River
Ruv 2	Hap 33	Hap 117	*H. tanzaniae*	Tanzania	Ruvu River
Ruv 3	Hap 34	-	*H. tanzaniae*	Tanzania	Ruvu River
Tan 1	Hap 29	Hap 106	*H. tanzaniae*	Tanzania	Rufiji River
Tan 2	Hap 30	Hap 107	*H. tanzaniae*	Tanzania	Rufiji River
Tan 3	Hap 30	Hap 108	*H. tanzaniae*	Tanzania	Rufiji River
Tan 4	-	Hap 109	*H. tanzaniae*	Tanzania	Rufiji River
Tan 5	Hap 31	Hap 108	*H. tanzaniae*	Tanzania	Rufiji River

Tan 6	-	Hap 110	*H. tanzaniae*	Tanzania	Rufiji River
Tan 7	-	Hap 111	*H. tanzaniae*	Tanzania	Rufiji River
Tan 8	-	Hap 111	*H. tanzaniae*	Tanzania	Rufiji River
Tan 9	-	Hap 108	*H. tanzaniae*	Tanzania	Rufiji River
Tan 10	Hap 30	Hap 112	*H. tanzaniae*	Tanzania	Rufiji River
Tan 11	-	Hap 107	*H. tanzaniae*	Tanzania	Rufiji River
Tan 12	-	Hap 107	*H. tanzaniae*	Tanzania	Rufiji River
Tan 13	Hap 30	Hap 113	*H. tanzaniae*	Tanzania	Rufiji River
Tan 14	-	Hap 114	*H. tanzaniae*	Tanzania	Rufiji River
Tan 15	-	Hap 107	*H. tanzaniae*	Tanzania	Rufiji River
Tan 16	-	Hap 107	*H. tanzaniae*	Tanzania	Rufiji River
Tan 17	Hap 32	Hap 107	*H. tanzaniae*	Tanzania	Rufiji River
Tan 18	-	Hap 110	*H. tanzaniae*	Tanzania	Rufiji River
Tan 19	-	Hap 107	*H. tanzaniae*	Tanzania	Rufiji River
Tan 20	-	Hap 115	*H. tanzaniae*	Tanzania	Rufiji River
Bu 1	Hap 13	Hap 25	*H. vittatus*	Mozambique	Buzi River
Bu 6	-	Hap 8	*H. vittatus*	Mozambique	Buzi River
Bu 10	-	Hap 26	*H. vittatus*	Mozambique	Buzi River
CK 1	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 2	Hap 8	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 4	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 6	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 7	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 8	Hap 8	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 9	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 10	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 13	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 15	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 17	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 18	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
CK 19	-	Hap 7	*H. vittatus*	South Africa	Inkomati River
CK 20	-	Hap 6	*H. vittatus*	South Africa	Inkomati River
Co 2	Hap 4	Hap 4	*H. vittatus*	Congo	Congo River
Co 3	Hap 3	Hap 3	*H. vittatus*	Congo	Congo River
Co 5	-	Hap 2	*H. vittatus*	Congo	Congo River
DRS 1	-	Hap 8	*H. vittatus*	Botswana	Okavango River
DRS 4	Hap 6	Hap 9	*H. vittatus*	Botswana	Okavango River
DRS 7	-	Hap 10	*H. vittatus*	Botswana	Okavango River

DRS 8	-	Hap 11	*H. vittatus*	Botswana	Okavango River
DRS 11	-	Hap 13	*H. vittatus*	Botswana	Okavango River
DRS 13	-	Hap 14	*H. vittatus*	Botswana	Okavango River
Gum 1	Hap 6	Hap 68	*H. vittatus*	Botswana	Okavango River
Gum 2	-	Hap 12	*H. vittatus*	Botswana	Okavango River
Gum 4	Hap 10	-	*H. vittatus*	Botswana	Okavango River
Gum 5	-	Hap 15	*H. vittatus*	Botswana	Okavango River
Gum 6	-	Hap 16	*H. vittatus*	Botswana	Okavango River
Gum 8	-	Hap 17	*H. vittatus*	Botswana	Okavango River
Hv 202	Hap 2	Hap 32	*H. vittatus*	Congo	Congo River
Hv 203	Hap 1	Hap 33	*H. vittatus*	Congo	Congo River
Hv 206	-	Hap 34	*H. vittatus*	Congo	Congo River
INGA 2	-	Hap 5	*H. vittatus*	Congo	Congo River
Kw 1a	-	Hap 52	*H. vittatus*	Congo	Kwango River
Kw 3a	Hap 5	Hap 53	*H. vittatus*	Congo	Kwango River
Kw 50	Hap 4	-	*H. vittatus*	Congo	Kwango River
Kw 214	Hap 4	-	*H. vittatus*	Congo	Kwango River
LZ 7	-	Hap 57	*H. vittatus*	Mozambique	Lower Zambezi River
LZ 8	-	Hap 58	*H. vittatus*	Mozambique	Lower Zambezi River
LZ 9	-	Hap 31	*H. vittatus*	Mozambique	Lower Zambezi River
Mz 1	-	Hap 27	*H. vittatus*	Zimbabwe	Lake Kariba
Mz 2	-	Hap 28	*H. vittatus*	Zimbabwe	Lake Kariba
Mz 4	-	Hap 29	*H. vittatus*	Zimbabwe	Lake Kariba
Mz 5	-	Hap 30	*H. vittatus*	Zimbabwe	Lake Kariba
Mz 7	-	Hap 28	*H. vittatus*	Zimbabwe	Lake Kariba
Mz 8	-	Hap 29	*H. vittatus*	Zimbabwe	Lake Kariba
Mz 9	-	Hap 28	*H. vittatus*	Zimbabwe	Lake Kariba
Mz 10	-	Hap 31	*H. vittatus*	Zimbabwe	Lake Kariba
Mz 11	-	Hap 27	*H. vittatus*	Zimbabwe	Lake Kariba
Mz 12	Hap 11	Hap 31	*H. vittatus*	Zimbabwe	Lake Kariba
Mz 14	Hap 8	Hap 27	*H. vittatus*	Zimbabwe	Lake Kariba
Nxa 1	-	Hap 8	*H. vittatus*	Botswana	Okavango River
Nxa 3	-	Hap 13	*H. vittatus*	Botswana	Okavango River
Nxa 9	Hap 6	-	*H. vittatus*	Botswana	Okavango River
Ok 4	-	Hap 18	*H. vittatus*	Botswana	Okavango Delta
Ok 16	-	Hap 8	*H. vittatus*	Botswana	Okavango River
Ok 17	-	Hap 19	*H. vittatus*	Botswana	Okavango River
Ok 18	-	Hap 8	*H. vittatus*	Botswana	Okavango River

Ok 19	-	Hap 20	*H. vittatus*	Botswana	Okavango River
Ok 20	-	Hap 21	*H. vittatus*	Botswana	Okavango River
Ok 21	-	Hap 22	*H. vittatus*	Botswana	Okavango River
Ok 22	-	Hap 23	*H. vittatus*	Botswana	Okavango River
Ok 23	-	Hap 24	*H. vittatus*	Botswana	Okavango River
Po 1	-	Hap 6	*H. vittatus*	South Africa	Pongola River
Po 2	-	Hap 7	*H. vittatus*	South Africa	Pongola River
Po 3	-	Hap 7	*H. vittatus*	South Africa	Pongola River
Po 4	-	Hap 59	*H. vittatus*	South Africa	Pongola River
Po 5	-	Hap 6	*H. vittatus*	South Africa	Pongola River
Po 6	-	Hap 7	*H. vittatus*	South Africa	Pongola River
Po 7	-	Hap 6	*H. vittatus*	South Africa	Pongola River
Po 8	-	Hap 6	*H. vittatus*	South Africa	Pongola River
Po 9	-	Hap 6	*H. vittatus*	South Africa	Pongola River
Po 11	-	Hap 6	*H. vittatus*	South Africa	Pongola River
Po 12	-	Hap 60	*H. vittatus*	South Africa	Pongola River
Po 14	-	Hap 6	*H. vittatus*	South Africa	Pongola River
Sh 1	-	Hap 31	*H. vittatus*	Malawi	Lower Shire River
Sh 2	Hap 9	Hap 54	*H. vittatus*	Malawi	Lower Shire River
Sh 3	-	Hap 30	*H. vittatus*	Malawi	Lower Shire River
Sh 4	-	Hap 55	*H. vittatus*	Malawi	Lower Shire River
Sh 5	-	Hap 31	*H. vittatus*	Malawi	Lower Shire River
Sh 6	-	Hap 56	*H. vittatus*	Malawi	Lower Shire River
Sh 7	Hap 8	Hap 56	*H. vittatus*	Malawi	Lower Shire River
Sh 8	-	Hap 30	*H. vittatus*	Malawi	Lower Shire River
SR 1	-	Hap 7	*H. vittatus*	Zimbabwe	Save River
SR 2	-	Hap 7	*H. vittatus*	Zimbabwe	Save River
SR 3	-	Hap 7	*H. vittatus*	Zimbabwe	Save River
SR 4	-	Hap 69	*H. vittatus*	Zimbabwe	Save River
Tang 1	Hap 12	-	*H. vittatus*	Zambia	Lake Tanganyika
TanLu 1	Hap 12	Hap 61	*H. vittatus*	Zambia	Lufubu River
TanLu 3	Hap 12	-	*H. vittatus*	Zambia	Lufubu River
TanLu 4	Hap 12	Hap 1	*H. vittatus*	Zambia	Lufubu River
TanLu 6	Hap 12	Hap 61	*H. vittatus*	Zambia	Lufubu River
TanLu 7	-	Hap 64	*H. vittatus*	Zambia	Lufubu River
TanLu 9	-	Hap 62	*H. vittatus*	Zambia	Lufubu River
TanLu 10	Hap 12	-	*H. vittatus*	Zambia	Lufubu River
TanLu 12	-	Hap 66	*H. vittatus*	Zambia	Lufubu River

TanLu 16	-	Hap 63	*H. vittatus*	Zambia	Lufubu River
TanLu 17	-	Hap 67	*H. vittatus*	Zambia	Lufubu River
TanLu 19	-	Hap 65	*H. vittatus*	Zambia	Lufubu River
TanLu 20	-	Hap 1	*H. vittatus*	Zambia	Lufubu River
TanLu 26	-	Hap 1	*H. vittatus*	Zambia	Lufubu River
Us 2	Hap 8	Hap 6	*H. vittatus*	Swaziland	Usuthu River
Us 3	Hap 8	Hap 6	*H. vittatus*	Swaziland	Usuthu River
UZ 50	Hap 6	Hap 8	*H. vittatus*	Zambia	Upper Zambezi River
UZ 51	-	Hap 35	*H. vittatus*	Zambia	Upper Zambezi River
UZ 52	-	Hap 36	*H. vittatus*	Zambia	Upper Zambezi River
UZ 53	-	Hap 37	*H. vittatus*	Zambia	Upper Zambezi River
UZ 54	-	Hap 8	*H. vittatus*	Zambia	Upper Zambezi River
UZ 55	-	Hap 38	*H. vittatus*	Zambia	Upper Zambezi River
UZ 56	Hap 6	Hap 37	*H. vittatus*	Zambia	Upper Zambezi River
UZ 57	-	Hap 23	*H. vittatus*	Zambia	Upper Zambezi River
UZ 58	-	Hap 39	*H. vittatus*	Zambia	Upper Zambezi River
UZ 59	-	Hap 8	*H. vittatus*	Zambia	Upper Zambezi River
UZ 60	-	Hap 40	*H. vittatus*	Zambia	Upper Zambezi River
UZ 61	Hap 6	Hap 41	*H. vittatus*	Zambia	Upper Zambezi River
UZ 62	-	Hap 42	*H. vittatus*	Zambia	Upper Zambezi River
UZ 63	-	Hap 38	*H. vittatus*	Zambia	Upper Zambezi River
UZ 64	-	Hap 43	*H. vittatus*	Zambia	Upper Zambezi River
UZ 66	-	Hap 44	*H. vittatus*	Zambia	Upper Zambezi River
UZ 67	-	Hap 45	*H. vittatus*	Zambia	Upper Zambezi River
UZ 68	-	Hap 46	*H. vittatus*	Zambia	Upper Zambezi River
UZ 69	-	Hap 44	*H. vittatus*	Zambia	Upper Zambezi River
UZ 70	Hap 6	Hap 8	*H. vittatus*	Zambia	Upper Zambezi River
UZ 71	Hap 7	Hap 38	*H. vittatus*	Zambia	Upper Zambezi River
UZ 72	-	Hap 36	*H. vittatus*	Zambia	Upper Zambezi River
UZ 73	-	Hap 47	*H. vittatus*	Zambia	Upper Zambezi River
UZ 75	-	Hap 48	*H. vittatus*	Zambia	Upper Zambezi River
UZ 77	-	Hap 8	*H. vittatus*	Zambia	Upper Zambezi River
UZ 78	-	Hap 49	*H. vittatus*	Zambia	Upper Zambezi River
UZ 79	-	Hap 50	*H. vittatus*	Zambia	Upper Zambezi River
UZ 80	-	Hap 8	*H. vittatus*	Zambia	Upper Zambezi River
UZ 82	-	Hap 37	*H. vittatus*	Zambia	Upper Zambezi River
UZ 83	-	Hap 23	*H. vittatus*	Zambia	Upper Zambezi River
UZ 84	-	Hap 51	*H. vittatus*	Zambia	Upper Zambezi River

Zam 1	-	Hap 28	*H. vittatus*	Mozambique	Lower Zambezi River
AY791404	Hap 14	-	*H. vittatus*	Congo	Congo River
Co 7	Hap 25	-	Group A	Congo	Congo River
Co 171	Hap 24	-	Group A	Congo	Congo River
Co 172	Hap 24	-	Group A	Congo	Congo River
Kw 4a	Hap 23	-	Group A	Congo	Kwango
Tshib 1	Hap 24	-	Group A	Congo	Lulua River
TanLu 5	Hap 42	Hap 70	Group B	Zambia	Lufubu River
TanLu 21	Hap 41	Hap 70	Group B	Zambia	Lufubu River
Bant 1	Hap 26	Hap 80	Group C	Zambia	Lake Bangweulu
Bant 2	Hap 26	Hap 81	Group C	Zambia	Lake Bangweulu
Cam 1	Hap 28	Hap 71	Group C	Cameroon	Dja River
Mw 1	-	Hap 72	Group C	Zambia	Luapula River
Mw 2	-	Hap 73	Group C	Zambia	Luapula River
Mw 3	-	Hap 74	Group C	Zambia	Luapula River
Mw 4	-	Hap 74	Group C	Zambia	Luapula River
Mw 5	Hap 27	Hap 75	Group C	Zambia	Luapula River
Mw 9	-	Hap 75	Group C	Zambia	Luapula River
Mw 16	Hap 26	Hap 76	Group C	Zambia	Luapula River
Mw 23	-	Hap 77	Group C	Zambia	Luapula River
Mw 28	-	Hap 72	Group C	Zambia	Luapula River
Mw 29	Hap 26	Hap 78	Group C	Zambia	Luapula River
Mw 30	-	Hap 79	Group C	Zambia	Luapula River
Ban 1	-	Hap 83	Group D	Zambia	Lake Bangweulu
Ban 3	-	Hap 88	Group D	Zambia	Lake Bangweulu
Ban 4	-	Hap 89	Group D	Zambia	Lake Bangweulu
Ban 5	-	Hap 88	Group D	Zambia	Lake Bangweulu
Ban 6	Hap 21	Hap 90	Group D	Zambia	Lake Bangweulu
Ban 7	-	Hap 88	Group D	Zambia	Lake Bangweulu
Ban 8	-	Hap 91	Group D	Zambia	Lake Bangweulu
Ban 9	-	Hap 88	Group D	Zambia	Lake Bangweulu
Ban 10	-	Hap 87	Group D	Zambia	Lake Bangweulu
Ban 11	-	Hap 92	Group D	Zambia	Lake Bangweulu
Ban 12	-	Hap 93	Group D	Zambia	Lake Bangweulu
Ban 13	Hap 21	Hap 94	Group D	Zambia	Lake Bangweulu
BanS 1	Hap 21	Hap 85	Group D	Zambia	Lake Bangweulu
BanS 3	Hap 21	Hap 86	Group D	Zambia	Lake Bangweulu
BanS 4	-	Hap 87	Group D	Zambia	Lake Bangweulu

BanS 10	-	Hap 95	Group D	Zambia	Lake Bangweulu
Cham 5	Hap 22	Hap 83	Group D	Zambia	Chambeshi River
Kasenga 1	Hap 21	Hap 82	Group D	Zambia	Luapula River
Kasenga 2	Hap 21	Hap 84	Group D	Zambia	Luapula River
Mw 6	Hap 26	Hap 91	Group D	Zambia	Luapula River
Mw 8	-	Hap 96	Group D	Zambia	Luapula River
Mw 11	-	Hap 91	Group D	Zambia	Luapula River
Mw 12	-	Hap 91	Group D	Zambia	Luapula River
Mw 13	-	Hap 83	Group D	Zambia	Luapula River
Mw 14	-	Hap 100	Group D	Zambia	Luapula River
Mw 15	-	Hap 97	Group D	Zambia	Luapula River
Mw 18	-	Hap 98	Group D	Zambia	Luapula River
Mw 19	Hap 21	Hap 99	Group D	Zambia	Luapula River
Mw 20	-	Hap 101	Group D	Zambia	Luapula River
Mw 21	-	Hap 91	Group D	Zambia	Luapula River
Mw 24	-	Hap 102	Group D	Zambia	Luapula River
Mw 25	-	Hap 104	Group D	Zambia	Luapula River
Mw 25a	Hap 21	Hap 103	Group D	Zambia	Luapula River
Mw 27	-	Hap 91	Group D	Zambia	Luapula River
Mw 31	-	Hap 105	Group D	Zambia	Luapula River

Appendix 2: Haplotype sequence alignment of *Hydrocynus cyt b* sequences

```
                10        20        30        40        50        60        70        80        90       100
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1    GTCCATATCTGTCGAGACGTAAACTATGGCTGAGCAGTCCGCAATATCCATGCTAACGGAGCATCCTTCTTCTTTATTTGTATTTATTTTCATATTGGCC
Hap2    ....................................................................................................
Hap3    ..A.................................................................................................
Hap4    ..................................................................................................T.
Hap5    ....................................................................................................
Hap6    ..................................................................................................T.
Hap7    ..................................................................................................T.
Hap8    ..................................................................................................T.
Hap9    ..................................................................................................T.
Hap10   ..................................................................................................T.
Hap11   ..................................................................................................T.
Hap12   ....................................................................................................
Hap13   ..................................................................................................T.
Hap14   ..A.................................................................................................
Hap15   .................T........C........TA....T.....T.....C.......................C..C........C..C..C....
Hap16   .................T........C........TA....T.....T.....C.......................C..C........C..C..C....
Hap17   .................T........C........TA....T.....T.....C.......................C..C........C..C..C....
Hap18   .................T..G...........G..TA....T.....T.....C.......................C..C.....C.....C..C....
Hap19   .................T..G...........G..TA....T.....T.....C.......................C..C.....C.....C..C....
Hap20   .................T..G...........G..TA....T.....T.....C.......................C..C.....C.....C..C....
Hap21   ............................................................................................C.....T.
Hap22   ............................................................................................C.....T.
Hap23   ..A........C........................A....A.....T..C..........................C..C...........C..C..T.
Hap24   ..A........C........................A....A.....T..C..........................C..C...........C..C..T.
Hap25   ..A........C........................A....A.....T..C..........................C..C...........C..C..T.
Hap26   ...............................................T..............................C..C.....C.....C..C..T.
Hap27   ...............................................T..............................C..C.....C.....C..C..T.
Hap28   ...............................................T..............................C..C.....C.....C..C..T.
Hap29   ........T.......................G..GA..........T.................T...........C..C...........C.......
Hap30   ........T.......................G..GA..........T.................T...........C..C...........C.......
Hap31   ........T.......................G..GA..........T.................T...........C..C...........C.......
Hap32   ........T.......................G..GA..........T.................T...........C..C...........C.......
Hap33   ........T.......................G..GA..........T.................T...........C..C...........C.......
Hap34   ........T.......................G..GA..........T.................T...........C..C...........C.......
Hap35   A....C.....C.....T........C.........A....A...........C.....G.....A..T........C.....C.....C..C..C....
Hap36   A....C.....C.....T........C.........A....A...........C.....G.....A..T........C.....C.....C..C..C....
Hap37   A....C.....C.....T........C.........A....A...........C.....G.....A..T........C.....C.....C..C..C....
Hap38   ...........C.....T........C....................T..............................C..C........C.....C....
Hap39   ...........C.....T..G.....C....................T..............................C..C........C.....C....
Hap40   ...........C.....T........C....................T..............................C..C........C.....C....
Hap41   ...............................................T....................T..T.....C........C.....C..C..T.
Hap42   ...............................................T....................T..T.....C........C.....C..C..T.
```

```
                110       120       130       140       150       160       170       180       190       200
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1    GAGGACTATACTATGGATCCTACCTCTACAAGGAAACATGAAATATCGGTGTAATCCTACTTCTACTAGTAATAATAACAGCATTCGTGGGCTACGTCCT
Hap2    ....................................................................................................
Hap3    ....................................................................................................
Hap4    ....................................................................................................
Hap5    ....................................................................................................
Hap6    ....................................................................................................
Hap7    ....................................................................................................
Hap8    ....................................................................................................
Hap9    ....................................................................................................
Hap10   .......G......................................T.....................................................
Hap11   ..............................................T.....................................................
Hap12   ....................................................................................................
Hap13   ....................................................................................................
Hap14   ....................................................................................................
Hap15   .......C.....C..C........T.....A.................C...G.....T.A..T...................................
Hap16   .......C.....C..C........T.....A.................C...G.....T.A..T...................................
Hap17   .......C.....C..C........T.....A.................C...G.....T.A..T...................................
Hap18   .......G.....C..C........T.......................C...G.....T.A..T.................G.................
Hap19   .......G.....C..C........T.......................C...G.....T.A..T.................G.................
Hap20   .......G.....C..C........T.......................C...G.....T.A..T.................G.................
Hap21   .G..G...........C.........................................G.........................................
Hap22   .G..G...........C.........................................G.........................................
Hap23   .............C...................................C...G....T.....C...................................
Hap24   .............C...................................C...G....T.....C...................................
Hap25   .............C...................................C...G....T.....C...................................
Hap26   .......C........G.........................................G.............................A...........
Hap27   .......C........G.........................................G.............................A...........
Hap28   .......C........G..............A..........................G.............................A...........
Hap29   ..........T....................A.....................G.G..G.....C...................................
Hap30   ..........T....................A.....................G.G..G.....C...................................
Hap31   ..........T....................A.....................G.G..G.....C...................................
Hap32   ..........T..C.................A.....................G.G..G.....C...................................
Hap33   ..........T....................A.....................G.G..G.....C...................................
Hap34   ..........T....................A.....................G.G..G.....C...................................
Hap35   ....C...........C..............A...........C.....C.....TT.......T...........G..C.....T..A...........
Hap36   ....C...........C..............A...........C.....C.....TT.......T...........G..C.....T..A...........
Hap37   ....C...........C..............A...........C.....C.....TT.......T...........G..C.....T..A...........
Hap38   ....GT.G.....C.....G...........A..............T..C...G.G..G.....T....................T..............
Hap39   ....GT.G.....C.....A...........A..............T..C...G.G..G.....T....................T..............
Hap40   ....GT.G.....C.....A...........A..............T..C...G.G..G.....T....................T..............
Hap41   ................G.........................................G.........................................
Hap42   ................G.........................................G.........................................
```

```
                210       220       230       240       250       260       270       280       290       300
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1    GCCTTGAGGACAGATGTCATTCTGAGGCGCAACAGTTATTACCAACCTACTCTCTGCAGTCCCCTACATTGGAGACACCCTAGTTCAATGAATTTGAGGC
Hap2    ....................................................................................................
Hap3    ....................................................................................................
Hap4    ....................................................................................................
Hap5    ....................................................................................................
Hap6    ....................................................................................................
Hap7    ....................................................................................................
Hap8    ....................................................................................................
Hap9    ....................................................................................................
Hap10   ....................................................................................................
Hap11   ....................................................................................................
Hap12   ....................................................................................................
Hap13   ....................................................................................................
Hap14   ....................................................................................................
Hap15   ..........................................T.....G..A..............T..C..............C........C......
Hap16   ..........................................T.....G..A..............T..C..............C........C.....T
Hap17   ..........................................T.....G..A..............T..C..............C........C.....T
Hap18   ..........................................T.....G..G..............T..C..............C........C..G..T
Hap19   ..........................................T.....G..G..............T..C..............C........C..G..T
Hap20   ..........................................T.....G..G..............T..C..............C........C..G..T
Hap21   ..........................................................A.............G........G..................
Hap22   ..........................................................A.............G........G..................
Hap23   ......G..........................C..............G..................G........G................C......
Hap24   ......G..........................C..............G..................G........G................C......
Hap25   ......G..........................C..............G...........................G................C......
Hap26   ................................................G....................C..G.......................G...
Hap27   ................................................G....................C..G.......................G...
Hap28   ................................................G....................C..G.......................G...
Hap29   C...........A................................................................................C......
Hap30   C...........A................................................................................C......
Hap31   C...........A................................................................................C......
Hap32   C...........A................................................................................C......
Hap33   C.....G.....A........................................................C.......................C......
Hap34   C.....G.....A........................................................C.......................C......
Hap35   ...C........A..A...........T.....C..C..C..A.....G.....C........A..TG.A..............C........C......
Hap36   ...C........A..A...........T.....C..C..C..A.....G.....C........A..TG.A..............C........C......
Hap37   ...C........A..A...........T.....C..C..C..A.....G.....C........A..TG.A..............CA.......C......
Hap38   ...............A..G....................C..............C..G..G.....T..C..T..T.....G...........C......
Hap39   ...............A..G....................C..............C..G..G.....T..C..T..T.....G...........C......
Hap40   ...............A..G....................C..............C..G..G.....T..C..T..T.....G...........C......
Hap41   ................................................G....................C..............................
Hap42   ................................................G....................C..............................
```

```
                310       320       330       340       350       360       370       380       390       400
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1    GGCTTCTCCGTTGACAACGCCACTTTAACTCGATTCTTCGCCTTCCATTTCCTACTTCCCTTCGCAATTGTAGCCGCCACCGCCCTACACGCCTTATTCC
Hap2    ...........................................................A........................................
Hap3    ...........................................................A........................................
Hap4    ....................................................................................................
Hap5    ...........................................................A........................................
Hap6    ....................................................................................................
Hap7    ....................................................................................................
Hap8    ....................................................................................................
Hap9    ....................................................................................................
Hap10   ....................................................................................................
Hap11   ....................................................................................................
Hap12   ....................................................................................................
Hap13   ....................................................................................................
Hap14   ...........................................................A........................................
Hap15   ........................C....C........T........C.....T.................G.....T........C........G..T.
Hap16   ........................C....C........T........C.....T.................G.....T........C........G..T.
Hap17   .......................CC....C........T........C.....T.................G.....T........C........G..T.
Hap18   .............................C........T........C..T..T.....A...........G..T..T........C...........T.
Hap19   .............................C........T........C..T..T.....A...........G..T..T........C...........T.
Hap20   .............................C........T........C..T..T.....A...........G..T..T........C...........T.
Hap21   ....................................................................................................
Hap22   ....................................................................................................
Hap23   ...............................................C.......................G.......................G....
Hap24   ...............................................C.......................G.......................G....
Hap25   ...............................................C.......................G.......................G....
Hap26   .....................................................G..............................................
Hap27   .....................................................G..............................................
Hap28   .....................................................G..............................................
Hap29   ......................................T........C............................T.........G.............
Hap30   ......................................T........C.......................................G.............
Hap31   ......................................T........C.......................................G.............
Hap32   ......................................T........C.......................................G.............
Hap33   ......................................T........C.......................................G.............
Hap34   ......................................T........C.......................................G.............
Hap35   ..A...........T..T..T..CC..........T..T........C...T..........T.....C..G..A..T.....T..T......C......
Hap36   ..A...........T..T..T..CC..........T..T........C...T..........T.....C..G..A..T.....T..T......C......
Hap37   ..A...........T..T..T..CC..........T..T........C...T..........T.....C..G..A..T.....T..T......C......
Hap38   ...........C...................................C...T.......A...........G........T.....C.............
Hap39   ...........C...................................C...T.......A...........G........T.....C.............
Hap40   ...........C...................................C...T.......A...........G........T.....C.............
Hap41   .......................................................................G............................
Hap42   .......................................................................G............................
```

```
                410       420       430       440       450       460       470       480       490       500
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1    TACATGAAACAGGCTCCAACAACCCCATTGGATTGAACTCAGACGCAGACAAAATCCCATTCCACCCA-TACTTCTCCTACAAAGACATCCTGGGCTTCC
Hap2    ....................................................................-...............................
Hap3    ....................................................................-...............................
Hap4    ....................................................................-...............................
Hap5    ..............................................G.....................-...............................
Hap6    ....................................................................-...............................
Hap7    ....................................................................-...............................
Hap8    ....................................................................-...............................
Hap9    ....................................................................-...............................
Hap10   ..........G.........................................................G...............................
Hap11   ....................................................................-...............................
Hap12   ....................................................................-...............................
Hap13   .G..................................................................-...............................
Hap14   ....................................................................-...............................
Hap15   .C.................T.....T..C..GC.A..T........GC..C.......G..T......-........A..............A......T
Hap16   .C.................T.....T..C..GC.A...........G...........G..T......-........A..............A......T
Hap17   .C.................T.....T..C..GC.A...........GA..........G..T......-........A..............A......T
Hap18   .C..C..G.................T..C..G..A...........G..............T......-........A.....................T
Hap19   .C..C..G.................T..C..G..A...........G..............T......-........A.....................T
Hap20   .C..C..G.................T..C..G..A...........G..............T......-........A.....................T
Hap21   ...............................G..............G.....................-.....T.................A..A....
Hap22   ...............................G..............G.....................-.....T.................A..A....
Hap23   .........................T..........................................-.......................A.......
Hap24   .........................T..........................................-.......................A.......
Hap25   .........................T..........................................-.......................A.......
Hap26   ..................................A.....G....................T.....G-...............................
Hap27   ..................................A.....G....................T......-...............................
Hap28   ..................................A.....G....................T.....G-...............................
Hap29   .G.....G..........................A...........G.....................-....................A..A..T....
Hap30   .G.....G..........................A...........G.....................-....................A..A..T....
Hap31   .G.....G..........................A...........G.....................-....................A..A..T....
Hap32   .G.....G..........................A...........G.....................-....................A..A..T....
Hap33   .G.....G..........................A...........G........T............-....................A.....T....
Hap34   .G.....G..........................A...........G........T............-....................A.....T....
Hap35   ....C..............T..............A....................T............-....................A..A..T...T
Hap36   ....C..............T..............A....................T............-....................A..A..T...T
Hap37   ....C..............T..............A....................T............-....................A..A..T...T
Hap38   .C.....G....................C..GC.A....................T............-.......................A.......
Hap39   .C.....G....................C..GC.A....................T............-.......................A.......
Hap40   .C.....G....................C..GC.A....................T............-.......................A.......
Hap41   .T.......................A....................G.....................-.....T....T..............T....
Hap42   .T................C......A..........................................-.....T....T..............T....
```

```
                510       520       530       540       550       560       570       580       590       600
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1    TAATCCTTCTGACTGGACTAACATCAC-TGGCCCTATTCTCCCCCAACCTTCTCGGGGACCCCGAGAATTTTACCCCCGCCAACCCCTTAGTTACACCCC
Hap2    ...........................-........................................................................
Hap3    ...........................-........................................................................
Hap4    ...........................-........................................................................
Hap5    ...........................-........................................................................
Hap6    ...........................-........................................................................
Hap7    ...........................-...............................................................A.........
Hap8    ...........................-........................................................................
Hap9    ...........................-........................................................................
Hap10   ...........................-........................................................................
Hap11   ...........................-........................................................................
Hap12   ...........................-........................................................................
Hap13   ...........................-........................................................................
Hap14   ...........................-........................................................................
Hap15   .......C..C..C.............-.......G..............C...........A.....C..C.....T..T.....T........G..A.
Hap16   .......C..C..C.............-.......G..............C...........A.....C..C.....T..T.....T........G..A.
Hap17   .......C..C..C.............-.......G..............C...........A.....C..C.....T..T.....T........G..A.
Hap18   .......C..T..C...........G.-.A..........................C.....A..A.....C.....T..T..............G..A.
Hap19   .......C..T..C...........G.-.A..........................C.....A..A.....C.....T..T..............G..A.
Hap20   .......C..T..C...........G.-.A..........................C.....A..A.....C.....T..T..............G..A.
Hap21   ..........T..............C.-.....................................A....................A.............
Hap22   ..........T..............C.C.....................................A....................A.............
Hap23   .G........T......T....G....-.A.............................................A...........T.............
Hap24   .G........T......T....G....-.A................................G..A.....................T.............
Hap25   .G........T......T....G....-.A................................G..A.....................T.............
Hap26   ....T.....T...........G..G.-..........T.....................................................G.......
Hap27   ....T.....T...........G..G.-..........T.....................................................G.......
Hap28   ....T.....T...........G..G.-..........T.....................................................G.......
Hap29   .G........T..C........T....-............................A.....A..A.....................C....C.......
Hap30   .G........T..C........T....-............................A.....A..A.....................C....C.......
Hap31   .G........T..C........T....-............................A.....A..A.....................C....C.......
Hap32   .G........T..C........T....-............................A.....A..A.....................C....C.......
Hap33   .G........T..C........T....-.A..........................A.....A..A.....................C....C.......
Hap34   .G........T..C........T....-.A..........................A.....A..A.....................C....C.......
Hap35   .......C..A....CC.....G..C.-.................A..T..C..G..A.....G..A..C..C.....T..T.....T........T..T.
Hap36   .......C..A....CC.....G..C.-.................A..T..C..G..A.....G..A..C..C.....T..T.....T........T..T.
Hap37   .......C..A....CC.....G..C.-.................A..T..C..G..A.....G..A..C..C.....T..T.....T........T..T.
Hap38   ....T..A..C.....G..G.......-.....G....................A........A..............T........A...........A.
Hap39   ....T..A..C.....G..G.......-.....G....................A........A..............T........A...........A.
Hap40   ....T..A..C.....G..G.......-.....G....................A........A..............T........A...........A.
Hap41   ....T.....T...........G....-......T....T..................................................G.....G....
Hap42   ....T.....T...........G....-......T....T..................................................G..........
```

```
                610       620       630       640       650       660       670       680       690       700
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1    CACATATTAAGCCAGAATGATACTTCCTATTCGCCTACGCAATCCTGCGCTCCATTCCAAACAAACTAGGTGGGGTTTTGGCCCTCCTCTTCTCTATTCT
Hap2    .G..............................................................................................T...........
Hap3    .G..............................................................................................T...........
Hap4    .G....................................................................C.................T...........
Hap5    .G..............................................................................................T...........
Hap6    .G..............................................................................................T...........
Hap7    .G..............................................................................................T...........
Hap8    .G..............................................................................................T...........
Hap9    .G..............................................................................................T...........
Hap10   .G..............................................................................................T...........
Hap11   .G..............................................................................................T...........
Hap12   .G..............................................................................................T...........
Hap13   .G..............................................................................................T...........
Hap14   .G..............................................................................................T...........
Hap15   ..............................................C.....T..C...........G..C..A..C..A.....T........A.....
Hap16   ..............................................C.....T..C...........G..C..A..C..A.....T........A.....
Hap17   ..............................................C.....T..C...........G..C..A..C..A.....T........A.....
Hap18   ..............................................C..A.....C...........G..C..A..C..A..T..T....C...A.....
Hap19   ..............................................C..A.....C...........G..C..A..C..A..T..T....C...A.....
Hap20   ..............................................C..A.....C...........G..G..A..C..A..T..T....C...A.....
Hap21   ..........................................................G..T.....G....................T...........
Hap22   ..........................................................G..T.....G....................T...........
Hap23   ..........................T...................C........C................................T...........
Hap24   ..........................T...................C........C................................T...........
Hap25   ..........................T...................C........C................................T...........
Hap26   .T.......................TT...................A..............T....................G.....T...........
Hap27   .T.......................TT...................A..............T....................G.....T...........
Hap28   .T.......................TT...................A..............T....................G.....T...........
Hap29   ..........A...................................C..............T..........................T...........
Hap30   ..........A...................................C..............T..........................T...........
Hap31   ..........A...................................C..............T..........................T...........
Hap32   ..........A...................................C..............T..........................T...........
Hap33   ..........A...................................C..............T..........................T...........
Hap34   ..............................................C..............T..........................T...........
Hap35   .C..C.....A..T.....G.....TT....T...........T..C........C..............C..A.....A..A..A..T.....C.....
Hap36   .C..C.....A........G.....TT....T...........T..C........C..............C..A.....A..A..A..T.....C.....
Hap37   .C..C.....A..T.....G.....TT....T...........T..C........C..............C..A.....A..A..A..T.....C.....
Hap38   .......C..............T.......................C........C.....T.....G..C..A.....A..A..T..A.....A.....
Hap39   .......C..............T.......................C........C.....T.....G..C..A.....A..A..T..A.....A.....
Hap40   .......C..............T.......................C........C.....T.....G..C..A.....A..A..T..A.....A.....
Hap41   .C....................................................................C..............T..T...........
Hap42   .C....................................................................C..............T..T...........
```

```
                710       720       730       740       750       760       770       780       790       800
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1    TGTCCTGATACTCGTTCCTATCCTCCACACATCAAAACAGCGAGCACTCACCTTCCGCCCCCTGTCCCAACTCCTATTCTGAACACTAATTGCAGATGTT
Hap2    ....................................................................................................
Hap3    ....................................................................................................
Hap4    ....................................................................................................
Hap5    ....................................................................................................
Hap6    ............A.......................................................................................
Hap7    ............A.......................................................................................
Hap8    ....T...............................................................................................
Hap9    .......................................A............................................................
Hap10   ............A.......................................................................................
Hap11   ....T...............................................................................................
Hap12   ....................................................................................................
Hap13   ....T...............................................................................................
Hap14   ..............................................................................T.....................
Hap15   C.....T.....T..C..C...........C........A........T...........T..A..T.................T...............
Hap16   C.....T.....T..C..C...........C........A........T...........T..A..T.................T...............
Hap17   C.....T.....T..C..C...........C........A........T...........T..A..T.................T...............
Hap18   ......A.....T......G..........C........A........T...........T..A..T..................T..............
Hap19   ......T.....T......G..........C........A........T...........T..A..T..................T..............
Hap20   ......T.....T......G..........C........A........T...........T..A..T..................T..............
Hap21   .......................................A..G..................T...................G..................
Hap22   .......................................A..G..................T...................G..................
Hap23   ......A..G....................T........A....................TT.A....................................
Hap24   ......A..G....................T........A....................TT.A....................................
Hap25   ......A..G....................T........A....................TT.A....................................
Hap26   .........G........C................................T.........T...........T.......G..................
Hap27   .........G........C................................T.........T...........T.......G..................
Hap28   .........G........C................................T.........T...........T.......G..................
Hap29   ...T..........................G........A..............T.............................G...............
Hap30   ...T..........................G........A..............T.............................G...............
Hap31   ...T...............G..........G........A..............T.............................G...............
Hap32   ...T..........................G........A..............T.............................G...............
Hap33   ...T...................................A..............T.............................G...............
Hap34   ...T...................................A..............T.............................G...............
Hap35   C.....T........C..C...T.A..T..C........A........T............T.A..............T.................C...
Hap36   C.....T........C..C...T.A..T..C........A........T............T.A..............T.................C...
Hap37   C.....T........C..C...T.A..T..C........A........T............T.A..............T.................C...
Hap38   ...A..T.....A.................C.................T............T.A..............T....................C
Hap39   ...A..T.....A.................C.................T............T.A..............T....................C
Hap40   ...A..T.....A.................C.................T............T.A..............T....................C
Hap41   ............T..............................A.................T...................G..G...............
Hap42   ............T..............................A.................T...................G..G...............
```

```
                  810       820
       ....|....|....|....|...
Hap1   GCCATTCTTACTTGAATTGGGGG
Hap2   .......................
Hap3   .......................
Hap4   .......................
Hap5   .......................
Hap6   .......................
Hap7   .......................
Hap8   .......................
Hap9   .......................
Hap10  .......................
Hap11  .......................
Hap12  .......................
Hap13  .......................
Hap14  .......................
Hap15  ..T..C..C..C........A..
Hap16  ..T..C..C..C........A..
Hap17  ..T..C..C..C........A..
Hap18  ..T..C.....C........A..
Hap19  ..T..C.....C........A..
Hap20  ..T..C.....C........A..
Hap21  ....................A..
Hap22  ....................A..
Hap23  .....C.....C........A..
Hap24  .....C.....C........A..
Hap25  .....C.....C........A..
Hap26  .....C..............A..
Hap27  .....C..............A..
Hap28  .....C..............A..
Hap29  ....................A..
Hap30  ....................A..
Hap31  ....................A..
Hap32  ....................A..
Hap33  ....................A..
Hap34  ....................A..
Hap35  .....C.....C.....C.....
Hap36  .....C.....C.....C.....
Hap37  .....C.....C.....C.....
Hap38  ..A........C........A..
Hap39  ..A........C........A..
Hap40  ..A........C........A..
Hap41  .......................
Hap42  .......................
```

Appendix 3: Additional data tables for Chapter 2.

Appendix 3.1: Parameters for the best-fit model (GTR+G) for *cyt b* sequence data.
G = Gamma.

	Cyt b
Model	GTR+G
Base frequencies:	
A Frequency	0.2529
C Frequency	0.3232
G Frequency	0.1481
T Frequency	0.2757
Rate Matrix:	
R(a) [A-C]	5.1501
R(b) [A-G]	62.5377
R(c) [A-T]	1.6271
R(d) [C-G]	2.6663
R(e) [C-T]	21.1383
R(f) [G-T]	1.0000
α	0.1691

Appendix 3.2: Kimura 2-parameter distances between *cyt b* lineages.
Values (%) below diagonal; standard errors (1000 replicates) above diagonal.

	H. vittatus	***H. forskahlii***	**D**	**A**	**C**	***H. tanzaniae***	***H. goliath***	***H. brevis***	**B**
H. vittatus	-	5.0	0.9	1.6	1.3	1.6	11.4	3.4	1.1
H. forskahlii	27.4	-	5.7	4.1	5.5	5.2	9.2	4.3	5.7
D	4.1	31.0	-	2.0	1.6	1.8	13.1	3.7	1.4
A	8.7	22.1	10.6	-	2.1	2.3	9.2	3.8	2.1
C	7.0	30.6	8.1	11.0	-	2.5	13.3	4.6	1.3
H. tanzaniae	9.0	27.4	10.3	11.6	13.5	-	12.1	4.5	2.3
H. goliath	48.3	44.0	55.2	41.1	55.6	52.5	-	9.8	12.5
H. brevis	20.6	24.1	21.7	22.2	26.2	24.5	44.8	-	4.6
B	5.4	29.9	7.2	11.5	6.9	12.9	53.3	25.3	-

Appendix 3.3: Kimura 2-parameter distances within *cyt b* lineages. Values with standard errors (SE) based on 1000 replicates.

Lineage	**Distance**	**SE**
H. vittatus	0.005	0.001
H. forskahlii	0.047	0.009
D	0.000	0.000
A	0.003	0.002
C	0.002	0.001
H. tanzaniae	0.006	0.002
H. goliath	0.002	0.001
H. brevis	0.002	0.001
B	0.004	0.002

Appendix 3.4: Table of lineage specific ΦST values.

	ΦST
H. vittatus	0.9749
H. goliath	0.97614
H. forskahlii	0.95562
H. brevis	0.97614
A	0.97567
D	0.97654
C	0.97621
H. tanzaniae	0.97425
B	0.97552

Appendix 4: Haplotype sequence alignment of Vittatus complex CR sequences

```
                10        20        30        40        50        60        70        80        90       100
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1    GTATTCTAAATTAAACTATTTTCTGCACTTGTACCCCGGACCTCGTGTCTCCAGGTGCATAATATGGTATATTATACATTATGGTTTA-GTACATAATAT
Hap2    ....................................................................................C...-...........
Hap3    ..................................................A.................................C...-...........
Hap4    ............................................A.A.........................................-...........
Hap5    ............................................A.A.................................C...-...........
Hap6    .............................................................G..........................-...........
Hap7    .............................................................G..........................-...........
Hap8    ............................................................................................-...........
Hap9    ........................................................................................-...........
Hap10   ........................................................................................-...........
Hap11   ........................................................................................-...........
Hap12   ........................................................................................-......G...
Hap13   ........................................................................................-...........
Hap14   ........................................................................................-...........
Hap15   ........................................................................................-...........
Hap16   ........................................................................................-...........
Hap17   ........................................................................................-...........
Hap18   ........................................................................................-...........
Hap19   .............................................................R..........................-...........
Hap20   ........................................................................................-...........
Hap21   ........................................................................................-...........
Hap22   .............................................................C..........................-...........
Hap23   ........................................................................................-...........
Hap24   ......................T......................................R..........................-...........
Hap25   .............................................................G..........................-...........
Hap26   .............................................................R..........................-...........
Hap27   .............................................................G..........................-...........
Hap28   ........................................................................................-...........
Hap29   .............................................................G..........................-...........
Hap30   .............................................................G..........................-...........
Hap31   .............................................................G..........................-...........
Hap32   ....................................................................................C...-...........
Hap33   ........................................................................................-...........
Hap34   ....................................................................................C...-...........
Hap35   ........................................................................................-...........
Hap36   ........................................................................................-...........
Hap37   ........................................................................................-...........
Hap38   ........................................................................................-...........
Hap39   ............................................A...........................................-...........
Hap40   ...............................C........................................................-...........
Hap41   ........................................................................................-...........
Hap42   ........................................................................................-...........
```

```
                10        20        30        40        50        60        70        80        90       100
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap43  ........................................................................................-...........
Hap44  ........................................................................................-...........
Hap45  ........................................................................................-...........
Hap46  ........................................................................................-...........
Hap47  ........................................................................................-...........
Hap48  ........................................................................................-...........
Hap49  ........................................................................................-...........
Hap50  ........................................................................................-...........
Hap51  ........................................................................................-...........
Hap52  ....................................................................................C...-...........
Hap53  ................................G....................A..............................C...-...........
Hap54  ........................................................................................-...........
Hap55  .............................................................G..........................-...........
Hap56  .............................................................G..........................-.......G...
Hap57  --...........................................................G..........................-...........
Hap58  ........................................................................................-...........
Hap59  .............................................................G..........................-...........
Hap60  .............................................................G..........................-...........
Hap61  ........................................................................................-...........
Hap62  ........................................................A...............................-...........
Hap63  ........................................................................................-...........
Hap64  ........................................................A...............................-...........
Hap65  ........................................................................................-...........
Hap66  ........................................................................................-...........
Hap67  ........................................................................................-...........
Hap68  ........................................................................................-...........
Hap69  .............................................................G..........................-...........
Hap70  .............................AA.GT...........C...C......................C......C.....C..G...........
Hap71  .............................C..GT...........CA..C..G...................C......C.......GAA......G...
Hap72  .............................C..GT...........CA..C..G...................C......C.......GAA......G...
Hap73  .............................C..GT...........CA..C..G...................C......C.......GAA......g...
Hap74  ......................T......C..GT...........CA..C..G...................C......C.......GAA......G...
Hap75  .............................C..GT...........CA..C..G...................C......C.......GAA......G...
Hap76  ......................T-..-..C..GT...........CA..C..G...................C......C.......GAA......G...
Hap77  ......................T-.....C..GT...........CA..C..G...................C......C.......GAA......G...
Hap78  ......................T-.....C..GT...........CA..C..G...................C......C.......GAA......G...
Hap79  ......................T......C..GT...........CA..C..G...................C......C.......GAA......G...
Hap80  .............................C..GT...........CA..C..G...................C......C.......GAA......G...
Hap81  .............................C.GGT...........CA..C..G...................C......C.......GAA......G...
Hap82  ..............................T..T...............C......................................-...........
Hap83  ..............................T..T...............C......................................-...........
Hap84  .........................A....T..T...............C......................................-...........
Hap85  ..............................T..T...............C......................................-...........
```

```
                   10        20        30        40        50        60        70        80        90       100
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap86      ..............................T..T...............C.......................................-...........
Hap87      ..............................T..T...............C.......................................-...........
Hap88      ..............................T..T...............C..................................C..-...........
Hap89      ..............................T..T...............C.......................................-...........
Hap90      ..............................T..T...............C.......................................-...........
Hap91      ..............................T..T...............C.......................................-...........
Hap92      ..............................T..T...............C.......................................-...........
Hap93      ..............................T..T...............C.......................................-...........
Hap94      ..............................T..T...............C.......................................-...........
Hap95      ..............................T..T...............C.......................................-.......G...
Hap96      ..............................T..T...............C.......................................-.......G...
Hap97      ..............................T..T...............C.......................................-...........
Hap98      ..............................T..T...............C.......................................-...........
Hap99      ..............................T..T...............C.......................................-...........
Hap100     ..............................T..T...............C.............C....T....................-........C..
Hap101     ......................T...-...T..T...............C.......................................-...........
Hap102     ..............................T..T...............C.................a....C................-...........
Hap103     ..............................T..T...............C.......................................-...........
Hap104     ..............................T..T...............C.......................................-...........
Hap105     ..............................T..T...............C.......................................-...........
Hap106     .........................AC..C.CG...T..G.....C.C.CTGG...................................-......T....
Hap107     .........................GC..C.CG...T..G.....C.C.CTGG...................................-......T....
Hap108     .........................GC..C.CG...T..G.....C.C.CTGG...................................-......T....
Hap109     .........................GC..C.CG...T..G.....C.C.CAGG...................................-......T....
Hap110     .........................GC..C.CG...T..G.....C.C.CTGG...................................-......T....
Hap111     .........................GC..C.CG...T..G.....C.C.CTGG...................................-......T....
Hap112     .........................GC..C.CG...T..G.....C.C.CTGG...................................-......T....
Hap113     ......................T..GC..C.CG...T..G.....C.C.CTGG...................................-......T....
Hap114     .........................GC..C.CG...T..G.....C.C.CTGG...................................-......T....
Hap115     .........................GC..C.CG...T..G.....C.C.CTGG...................................-......T....
Hap116     .............................C.C....G..G.....CAC.C......................................-......T....
Hap117     .............................C.C....G..G.....CAC.C......................................-......T....
```

```
                110       120       130       140       150       160       170       180       190       200
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1    GGTAA--CATAATTTTACATAATGGTTTTAAAACATTAAACTGAAAGTACCATAAAACATATTAAACACATTCAGGAATAATACCAAACTAAGACAAACA
Hap2    ....---.............G.......G.......................................................................
Hap3    ....---.............G.......G.......................................................................
Hap4    .------.............G.......A.......................................................................
Hap5    ....---.............G.......A.......................................................................
Hap6    .------.....................G.............A....C................T.........A.........................
Hap7    .------.....................G.............A....C....................................................
Hap8    .------.....................A.............A.....................................TG..................
Hap9    .------.....................A.............A.....................................TG.......G......
Hap10   .....--.....................A.............A.....................................TG..................
Hap11   ....---.....................A.............A.....................................TG..................
Hap12   .------.....................A.............A.....................................TG..................
Hap13   .------.....................A.............A.....................................T...................
Hap14   .------.....................A.............A.....................................TG..................
Hap15   .------.....................A.............A.....................................TG.......G......
Hap16   .------.....................A.............A.....................................TG..................
Hap17   .------.....................A.............A...A.................................TG..................
Hap18   .------.....................A.............A.....................................TG..................
Hap19   .------.....................A.............A....Y................................YR..................
Hap20   .------.....................A.............A....c.................................G..................
Hap21   .------.....................A.............A.....................................TG..................
Hap22   .------.....................A.............A....C.................................N..................
Hap23   .------.....................A.............A.....................................TG..................
Hap24   .------.....................A.............A....Y................................YR..................
Hap25   .------.....................A.............A....M....................................................
Hap26   .------.....................R.............A....C................................YR..................
Hap27   .------.....................A.............A....C....................................................
Hap28   .------.....................A.............A....C....................................................
Hap29   .------.....C...............A.............A....C....................................................
Hap30   .------.....................A........G....A....C....................................................
Hap31   .------.....................A.............A....C....................................................
Hap32   ..C.---.............G.......A.............AG....................................T...................
Hap33   ....---...........................................................................G.........G......
Hap34   ....---.............G.......A........G..........................................T...................
Hap35   .------.....................A.............A................C.....G..............TG..................
Hap36   .------.....................A.............A..................................C..TG..................
Hap37   .------.....................A.............A..................................C..TG..................
Hap38   .------.....................A.............A.....................................TG..................
Hap39   .------.....................A.............A..................................C..TG..................
Hap40   .------.....................A.............A..G..................................TG..................
Hap41   .------.....................A.............A..G..................................TG..................
Hap42   .------.....................A.............A.G...................................TG..................
```

```
               110       120       130       140       150       160       170       180       190       200
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap43  .------.................A...........A..........................................TG..............
Hap44  .------.................A...........A..........................................TG..............
Hap45  .------.................A...........A..........................................TG..............
Hap46  .------.................A...........A..........................................TG..............
Hap47  .------.................A...........A..........................................TG..............
Hap48  .------.................A...........A..........................................TG..............
Hap49  .------.................A...........A..........................................TG..............
Hap50  .------.................A...........A..........................................TG..............
Hap51  .------.................A...........A..........................................TG..............
Hap52  ....---.............G.......A...........AG.........................................T...............
Hap53  ....---.............G.......A........G.............................................T...............
Hap54  .------.................G...........A...............................A...................A.....
Hap55  .------.................A...........A....C.....................................................
Hap56  .------.................A...........A....C.....................................................
Hap57  .------.................A........Y..AA....C....................................................
Hap58  .------.................A...........A....C........................M.A.........................
Hap59  .------.................G...........A....C................T.........A.........................
Hap60  .------.................G...........A....C................T.........A.........................
Hap61  ....---.....................C.......................................................................
Hap62  ....R--.....................M...................................................................A.......
Hap63  .....AT.............................................................................................R.......
Hap64  .....G-.....................C.......................................................................
Hap65  .....A-.............................................................................................R.......
Hap66  ....---.............G...............................................................................
Hap67  ....---.............................................................................................
Hap68  .------.................A...........A..........................................TG..............
Hap69  .------.................G...........A....C.....................................................
Hap70  ....---.....................A...........T.A.G.....T.......................A..G.....TTT.....G.GT.....
Hap71  ....---.............G.......A........C..T.A.G..........G..................A........TTG.....G.GT.....
Hap72  ....---.....................A........C.GT.A.G.............................A........TT......G.GT..G..
Hap73  ....---.....................A........C.GT.A.g.......................................TT......G.GT..G..
Hap74  ....---.....................A........C.GT.A.GG............................A........TT......G.GT..G..
Hap75  ....---.....................A........C.GT.A.G.............................A........TT......G.GT..G..
Hap76  ....---.....................A........C.GT.A.G.............................A........TT......G.GT..G..
Hap77  ....---.....................A........C.GT.A.G.......................................TT......G.GT..G..
Hap78  ....---.....................A........C.GT.A.G.......................................TT......G.GT..G..
Hap79  ....---.....................A........C.GT.A.G.............................A........TT......G.GT..G..
Hap80  ....---.....................A........C.GT.AG..............................A........TT......G.GT..G..
Hap81  ....---.....................A........C.GT.A.G.............................A........TT......G.GT..G..
Hap82  .....AA.....................A...........ACA......................G.....C..A.........T.........GT.....
Hap83  .------.................A...........ACA....C..................G.....C..A......C..T.........GT.....
Hap84  ....---.....................A...........ACA....C..................G.....C..A.........T.........GT.....
Hap85  .------.................A...........ACA....C..................G........A.........T.........GT.....
Hap86  .------.................A...........ACA....C..................G.....C..A.........T.........GT.....
```

```
                110       120       130       140       150       160       170       180       190       200
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap87   ....---.....................A...........ACA....C.................G.....C..A.........T........GT.....
Hap88   ....---.....................A...........ACA....C.................G.....C..A.........T........GT.....
Hap89   ....---.....................A...........ACA....C.......................C..A.........T........GT.....
Hap90   .....--.....................A...........ACA....C.................G.....C..A.........T........GT.....
Hap91   .------.....................A...........ACA....C.................G.....C..A.........T........GT.....
Hap92   .------.....................A...........ACA....C.......................C..A.........T........GT.G...
Hap93   ....---.....................A...........ACA....C.................G.....C..A.........T........GT.....
Hap94   ....---.............G.......A...........ACA....C.................G.....C..A.........T......G.GT.....
Hap95   .------.....................A...........ACA....C.......................C..A.........T........GT.....
Hap96   ....---.....................A...........ACA....C.................G.....C..A.........T........GT.....
Hap97   ....---.....................A...........ACA....C.......................C..A.........T........GT.....
Hap98   ....---.....................A...........ACA....C.................G........A.........T........GT.....
Hap99   .------.....................A...........ACA...A........................C..A.........T........GT.....
Hap100  ....---.............G.......A...........ACA....C.................G.....C..A.........T........GT.....
Hap101  .-------....................A...........ACA....C.......................C..A.........T........GT.....
Hap102  .A..---.....................A...........ACA....C.................G.....C..A..................GT.....
Hap103  .------.....................A...........ACA....C.................G.....C..A.........T........GT.....
Hap104  .------.....................A...........ACA....C.................G.....C..A.........T........GT.....
Hap105  .------.....................A...........ACA....C.................G.....C..A.........T........GT.....
Hap106  ...T---......A..............A..G.....C..T.A.GTA......TC....C.C...GT.......A..G.....AA.-.T....GT.....
Hap107  ...T---......A..............A..G.....C..T.A.GTA......TC....C.C...GT.......A..G.....AA.-.T....GT..G..
Hap108  ...T---......A..............A..G.....C..T.A.GTA......TC....C.C...GT.......A..G.....AA.-.T....GT.....
Hap109  ...----......A..............A..G.....C..T.A.GTA......TC....C.C...GT.......A..G.....AA.-.T....GT..G..
Hap110  ...T---......A..............A..G.....C..T.A.GTA......TC....C.C...GT.......A..G.....AA.-.TC...GT..G..
Hap111  ...T---......A..............A..G.....C..T.A.GTA......TC....C.C...GT.......A..G.....AA.-.T....GT..G..
Hap112  ...T---......A..............A..G.....C..T.A.GTA......TC....C.C...GT.......A..G.....AA.-.T....GT.....
Hap113  ...T---......A..............A..G.....C..T.A.GTA......TC....C.C...GT..........G.....AA.-.TC...GT..G..
Hap114  ...T---......A..............A..G.....C..T.A.GTA......TC....C.C...GT.......A..G.....AA.-.T....GT..G..
Hap115  ...T---......A..............A..G.....C..T.A.GTA......TC....C.C...GT.......A..G.....AA.-.T....GT..G..
Hap116  ...T---......A..............A..G.....C..T.A.GTA......TT....C.C...GT.......A..G.....AAG-......GT.....
Hap117  ...T---......A..............A..G.....C..T.A.GTA......TT....C.C...GT.......A..G.....AA.-......GT.....
```

```
                  210       220       230       240       250       260       270       280       290       300
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap1      TAAAGCATCCTATTAGTTACTACAGAAAAAATAATTAAACCTGATAAATTGAATAATCCCCATAACTCCAATAAAACATTTTCCTTGTAAAACACAACCA
Hap2      ....................................G...............................................................
Hap3      ....................................G...............................................................
Hap4      ....................................................................................................
Hap5      ....................................................................................................
Hap6      ...........................................G........................................................
Hap7      ....................................................................................................
Hap8      ....................................................................................................
Hap9      ............................G.......G...............................................................
Hap10     ............................G.......G...............................................................
Hap11     ............................G.......G...............................................................
Hap12     ............................G.......G...............................................................
Hap13     ....................................G...............................................................
Hap14     ............................G.......G...................C...........................................
Hap15     ............................GG......G...............................................................
Hap16     ....................................G...............................................................
Hap17     ....................................G.......................................................T.......
Hap18     ............................G.......G...............................................................
Hap19     ....................................R...............................................................
Hap20     ............................g.......G...............................................................
Hap21     ....................................G...............................................................
Hap22     ...........................................N........................................................
Hap23     ............................N.......G...............................................................
Hap24     ....................................R...............................................................
Hap25     ..........................................................................................G.........
Hap26     ............................R.......R...............................................................
Hap27     ....................................................................................................
Hap28     ....................................................................................................
Hap29     ....................................................................................................
Hap30     ....................................................................................................
Hap31     ....................................................................................................
Hap32     ................................R............................................................G......
Hap33     ...................................G.........................G.....G................................
Hap34     ........................................................................T...........................
Hap35     ............................G.......G...............................................................
Hap36     ....................................G...............................................................
Hap37     ............................G.......G...............................................................
Hap38     ....................................G...............................................................
Hap39     ............................G.......G...............................................................
Hap40     ............................G.......G...............................................................
Hap41     ............................G.......G...............................................................
Hap42     ............................G.......G...............................................................
```

```
               210       220       230       240       250       260       270       280       290       300
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap43   ....................................G.......................................................T.......
Hap44   ............................G.......G...............................................................
Hap45   ............................G.......G............................T..................................
Hap46   ....................................G.......................................................T.......
Hap47   ............................G.......G...............................................................
Hap48   ............................G.......G...............................................................
Hap49   ............................G.......G...............................................................
Hap50   ............................G.......G...............................................................
Hap51   ....................................G.......................................................T.......
Hap52   ....................................................................................................
Hap53   ........................................................................T...........................
Hap54   ....................................................................................................
Hap55   ........................................................................G...........................
Hap56   ....................................................................................................
Hap57   ........................T..............................................C..............T.............
Hap58   ....................................................................................................
Hap59   ...........................................G........................................................
Hap60   ....................................................................................................
Hap61   ....................................................................................................
Hap62   ....................................................................................................
Hap63   ....................................................................................................
Hap64   ........................R...........................................................................
Hap65   ....................................................................................................
Hap66   ........................................................................G...........................
Hap67   ........................................................................G...........................
Hap68   ............................G.......G.............................................................T.
Hap69   ....................................................................................................
Hap70   ...............................C..C.G............C.....................................C...G........
Hap71   ........T.......C.................C..............C......................G..............C...GA.......
Hap72   ........T.......C.................C..............C.....................................C..GGA....TT.
Hap73   ........T.......C.................C..............c.....................................C..GGA.....T.
Hap74   ........T.......C.................C..............C.....................................C..GGA.....T.
Hap75   ........T.......C.................C..............C.....................................C..GGA.....T.
Hap76   ........T.......C.................C..............C.....................................C..GGA.....T.
Hap77   ........T.......C.................C..............C.....................................C..GGA.....T.
Hap78   ........T.......C.................C..............C.....................................C..GGA.....T.
Hap79   ........T.......C.................C..............C.....................................C..G.A.....T.
Hap80   ........T.......C.................C..............C.....................................C..GGA.....T.
Hap81   ........T.......C.................C..............C.....................................C..GGA.....T.
Hap82   ........................a.........C.....................................G..............C..G.T.......
Hap83   ..................................C......................T..............G..............C....T.......
Hap84   ..................................C......................T..............G..............C....T.......
Hap85   ..................................C.G....................T..............G..............C....T.......
Hap86   ..................................C......................T..............G..............C....T.......
```

```
                210       220       230       240       250       260       270       280       290       300
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hap87   ..................................C......................T..............G..............C....T.......
Hap88   ..................................C......................T..............G..............C....T.......
Hap89   ..................................C......................T..............G..............C....T.......
Hap90   ..................................C......................T..............G..............C....T.......
Hap91   ..................................CC.....................T..............G..............C....T.......
Hap92   ..................................C......................T..............G..............C....T.......
Hap93   ..................................C......................T..............G..............C....T.......
Hap94   ..................................C......................T..............G..............C....T.......
Hap95   ..................................C......................T..............G..............C....T.......
Hap96   ..................................C......................T.............................C....T.......
Hap97   ..................................C......................T..............G..............C....T.......
Hap98   ..................................C......................T..............G..............C....T.......
Hap99   ..................................C......................T..............G..............C....T.......
Hap100  ..................................C......................T..............G..............C....T.......
Hap101  ..................................C......................T..............G..............C....T.......
Hap102  ..................................C......................T..............G..............C....T.......
Hap103  ..................................C......................T..............G..............C..G.T.......
Hap104  ..................................C......................T..............G..............C....T.......
Hap105  ..................................C......................T..............G..............C....T.......
Hap106  ........TAA.....C.......A......C....G............C....................TC...............A..T.......T.
Hap107  ........TAA.....C.......A......C....G............C....................TC...............A..T.......T.
Hap108  ........TAA.....C.......A......C....G............C....................TC...............A..T.......T.
Hap109  ........TAA.....C.......A......C....G............C....................TC...............A..T.......T.
Hap110  ........TAA.....C.......A......C....G............C....................TC...............A..T.......T.
Hap111  ........TAAG....C.......A......C....G............C....................TC...............A..T.......T.
Hap112  ........TAA.....C.......A......C....G............C....................TCG..............A..T.......T.
Hap113  ........TAA.....C.......A......C....G............C....................TC...............A..T.......T.
Hap114  ........TAA.....C.......A......C....GG...........C....................TC...............A..T.......T.
Hap115  ........AAA.....C.......A......C....G............C....................TC...............A..T.......T.
Hap116  ........TAA.....C.......A........................C....................TC...............A..TCT.....T.
Hap117  ........TAA.....C.......A........................C....................TC...............A..TCT.....T.
```

```
                310       320       330       340       350       360
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....
Hap1   AAATCGGACCTCAAAATATTAATGTAGTAAGAGACCACCAACCGGTT-ATTTCTAAATGTTAATAATCC
Hap2   ...............................................-.....................
Hap3   ...............................................-.....................
Hap4   ...............................................-.....................
Hap5   ...............................................-.....................
Hap6   ....T.....................................T....-.....................
Hap7   ....T.....................................T....-.....................
Hap8   ...............................................-...................T.
Hap9   ........T......................................-...................T.
Hap10  ...............................................-...................T.
Hap11  ...............................................-...................T.
Hap12  ...............................................-...................T.
Hap13  ....................................C..........-...................T.
Hap14  ...............................................-...................T.
Hap15  ........T......................................-...................T.
Hap16  ...............................................-...........C.......T.
Hap17  ...............................................-...................T.
Hap18  ...............................................-...................T.
Hap19  ....T..........................................-...................Y.
Hap20  ....T..........................................-..............-......
Hap21  ...............................................-..................AT.
Hap22  ....T..........................................-......N..............
Hap23  ...............................................-...................T.
Hap24  ....T..........................................-...................Y.
Hap25  ....T..........................................-.....................
Hap26  ....Y......................................A...-...................T.
Hap27  ....T..........................................-......G..............
Hap28  ....T..........................................-.....................
Hap29  ....T..........................................-.....................
Hap30  ....T..........................................-.....................
Hap31  ....T..........................................-.....................
Hap32  ...............................................-.....................
Hap33  ...........................................A...-.....................
Hap34  ...............................................-T....................
Hap35  ...............................................-...................T.
Hap36  ...............................................-...................T.
Hap37  ...............................................-...................T.
Hap38  ...............................................-...................T.
Hap39  ...............................................-...................T.
Hap40  ..........................A...................--...................T.
Hap41  ..........................A...................--...................T.
Hap42  ...............................................-...................T.
```

```
               310       320       330       340       350       360
       ....|....|....|....|....|....|....|....|....|....|....|....|....|....
Hap43  ........................................C..C-..-.................T.
Hap44  ............................................-..-.................T.
Hap45  ...............................................-.................T.
Hap46  ...............................................-.................T.
Hap47  ........................................C...-..-.................T.
Hap48  ........................................T...-..-.................T.
Hap49  ........................................T......-.................T.
Hap50  .......................................CT...-..-.............C...T.
Hap51  .......................................C....-..-.................T.
Hap52  .......................................TC......-...................
Hap53  ........................................T......-...................
Hap54  ...............................................-...................
Hap55  ....T..........................................-...................
Hap56  ....T..........................................-...................
Hap57  ....T..................................C.......-...............--..
Hap58  ....T......................................A...-......G............
Hap59  ....T..................................M..T....-...................
Hap60  ....T.....................................T....-...................
Hap61  ...............................................-...................
Hap62  ...............................................-...................
Hap63  ......A........................................-...................
Hap64  ...............................................-...................
Hap65  ......A........................................-...................
Hap66  ...............................................-...................
Hap67  ...............................................-...................
Hap68  ........................................T......-.................T.
Hap69  ....T............................T........T....-...................
Hap70  ..........................................T....T..AA..T..C.C....CC.T.
Hap71  ..................................T.......T....T..AA.....C.C...CT..T.
Hap72  ..................................T.......T....T..AA.....C.C...CT..T.
Hap73  ..................................T.......T....T..AA.....C.C...CT..T.
Hap74  ..................................t.......T....T..AA.....C.C...CT..T.
Hap75  ..................................T.......T....T..AA.....C.C...CT..T.
Hap76  ..................................T.......T....T..AA.....C.C...CT..T.
Hap77  ..................................T.......T....T..AA.....C.C....T..T.
Hap78  ..................................T.......T....T..AA.....C.C...CT..T.
Hap79  ..................................T.......T....T..AA.....C.C...CT..T.
Hap80  ..................................T.......T....T..AA.....C.C...CT..T.
Hap81  ..................................T.......T....T..AA.....C.C...CT..T.
Hap82  ....T.............A.......................T....T..AC..C..............
Hap83  ....T.............A.......................T....T..AC..C..............
Hap84  ....T.............A.......................T....T..AC..C..............
Hap85  ....T.............A.......................T....T..AC..C..............
Hap86  ....T.............A.......................T....T..AC..C..............
```

```
                310       320       330       340       350       360
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....
Hap87   ....T.............A.......................T....T..AC..C..............
Hap88   ....T.............A.......................T....T..AC..C..............
Hap89   ....T.............A.......................T....T..AC..C..............
Hap90   ....T.............A.......................T....T..AC..C..............
Hap91   ....T.............A.......................T....T..AC..C..............
Hap92   ....T.............A.......................T....T..AC..C..............
Hap93   ....T.............A.............R.........T....T..AC..C.........G....
Hap94   ....T.............A.......................T....T..AC..C..............
Hap95   ....T.............A.......................T....T..AC..C..............
Hap96   ....T.............A.......................T....T..AC..C..............
Hap97   ....T.............A.......T...............T....T..AC..C..............
Hap98   ....T.............A.......................T....T..AC..C..............
Hap99   ....T.............A.......................T....T..AC..C..............
Hap100  ....T.............A.......................T....T..AC..C..............
Hap101  ....T.............A.......................T....T..AC..C..............
Hap102  ....T.............A.......................T....T..AC..C..............
Hap103  ....T.............A.......................T....T..AC..C..............
Hap104  ....T.............A.......................T....T..AC..C..............
Hap105  ....T.............A.......................T....T..AC..C..............
Hap106  ....T.....................................T....T..A......C.C....C....
Hap107  ....T.....................................T....T..A......C.C....C....
Hap108  ....T.....................................T....T..A......C.C....C....
Hap109  ....T.....................................T....T..G......C.C....C....
Hap110  ....T.....................................T....T..A......C.C....C....
Hap111  ....T.....................................T....T..A......C.C....C....
Hap112  ....T.....................................T....T..A......C.C....C....
Hap113  ....T.....................................T....T..A......C.C....C....
Hap114  ....T.....................................T....T..A...T..C.C....C....
Hap115  ....T.....................................T....T..A...T..C.C....C....
Hap116  ....T........G............................T....T..A......C.C....C....
Hap117  ....T........G............................T....T..A......C.C....C....
```

Appendix 5: Additional data tables for Chapter 3

Appendix 5.1: Parameters for the best-fit model for the CR sequence data

G = Gamma.

	CR Vittatus complex	**CR *H. vittatus***
Model	HKY+G	TrN+G
Base frequencies:		
A Frequency	0.3998	0.3892
C Frequency	0.1769	0.1927
G Frequency	0.1368	0.1158
T Frequency	0.2865	0.3023
Rate Matrix:		
R(a) [A-C]	-	1.0000
R(b) [A-G]	-	12.0094
R(c) [A-T]	-	1.0000
R(d) [C-G]	-	1.0000
R(e) [C-T]	-	5.6067
R(f) [G-T]	-	1.0000
α	0.3348	0.2498

Appendix 5.2: Kimura 2-parameter distances between CR lineages

Kimura 2-parameter values as percentages are shown below the diagonal; standard errors (based on 1000 bootstrap replicates) are shown above the diagonal.

	H. vittatus	**B**	**C**	**D**	***H. tanzaniae***
H. vittatus	-	3.1	4.6	1.9	5.7
B	13.1	-	2.1	2.6	4.1
C	19.5	8.7	-	3.7	1.7
D	7.7	11.0	17.0	-	5.8
H. tanzaniae	25.0	19.7	22.5	26.9	-

Appendix 5.3: Kimura 2-parameter distances within CR lineages

Kimura 2-parameter values as percentages; standard errors are based on 1000 bootstrap replicates.

Lineage	**Distance**	**SE**
H. vittatus	1.0	0.2
B	-	-
C	1.1	0.3
D	0.8	0.2
H. tanzaniae	2.1	0.5

Appendix 5.4: Kimura 2-parameter distances between *H. vittatus* populations

Kimura 2-parameter values as percentages are shown below the diagonal; standard errors (based on 1000 bootstrap replicates) are shown above the diagonal.

	Lufubu	**Congo**	**Crocodile**	**Okavango**	**Buzi**	**Usutu**	**Middle Zambezi**	**Upper Zambezi**	**Kwango**	**Shire**	**Lower Zambezi**	**Pongola**	**Save**
Lufubu	-	0.4	0.6	0.3	0.4	0.7	0.3	0.3	0.7	0.3	0.4	0.6	0.5
Congo	1.4	-	0.8	0.5	0.5	0.8	0.5	0.5	0.5	0.5	0.5	0.8	0.6
Crocodile	1.4	2.4	-	0.5	0.5	0.0	0.5	0.5	0.9	0.5	0.5	0.1	0.4
Okavango	0.7	1.7	1.1	-	0.2	0.5	0.1	0.1	0.7	0.1	0.2	0.5	0.3
Buzi	0.7	1.6	1.1	0.4	-	0.6	0.1	0.2	0.7	0.2	0.2	0.5	0.4
Usutu	1.5	2.5	0.0	1.1	1.1	-	0.5	0.6	1.0	0.5	0.5	0.1	0.4
Middle Zambezi	0.6	1.5	0.9	0.3	0.3	1.0	-	0.1	0.7	0.1	0.2	0.5	0.3
Upper Zambezi	0.8	1.7	1.1	0.5	0.5	1.2	0.3	-	0.7	0.1	0.2	0.5	0.3
Kwango	1.9	1.9	2.7	2.0	2.0	2.8	1.8	2.0	-	0.7	0.7	0.9	0.8
Shire	0.7	1.7	1.0	0.4	0.4	1.0	0.2	0.4	1.9	-	0.2	0.5	0.3
Lower Zambezi	0.9	1.8	1.1	0.6	0.5	1.2	0.4	0.6	2.2	0.6	-	0.5	0.4
Pongola	1.3	2.3	0.2	1.0	1.0	0.2	0.8	1.0	2.6	0.9	1.0	-	0.3
Save	0.9	1.9	0.6	0.6	0.6	0.7	0.4	0.6	2.2	0.6	0.8	0.5	-

Appendix 5.5: Kimura 2-parameter distances within *H. vittatus* populations for the CR data set

Kimura 2-parameter values as percentages; standard errors (SE) are based on 1000 bootstrap replicates.

Lineage	Distance	SE	Lineage	Distance	SE
Lufubu	0.4	0.2	**Upper Zambezi**	0.5	0.2
Congo	1.7	0.4	**Kwango**	1.9	0.8
Crocodile	0.1	0.1	**Shire**	0.4	0.2
Okavango	0.4	0.1	**Lower Zambezi**	0.8	0.3
Buzi	0.4	0.3	**Pongola**	0.2	0.2
Usutu	0.0	0.0	**Save**	0.2	0.1
Middle Zambezi	0.1	0.1			

Appendix 5.6: Table of lineage specific ΦST values

	ΦST
H. vittatus	0.91001
D	0.91226
B	N/A
H. tanzaniae	0.91136
C	0.91197

Appendix 5.7: Gsi values calculated to quantify the degree of shared evolutionary history between populations of *H. vittatus*

P-values calculated based on a permutation test (10 000 replicates)

Population	**GSI value**	**P-value**
Buzi	0.1584	< 0.030
Okavango	0.3726	< 0.001
Upper Zambezi	0.6089	< 0.001
Pongola	0.4079	< 0.001
Usutu	0.1625	< 0.050
Inkomati	0.4829	< 0.001
Save	0.3940	< 0.001
Shire	0.3908	< 0.001
Middle Zambezi	0.4055	< 0.001
Lower Zambezi	0.2192	< 0.030
Congo	0.5878	< 0.001
Kwango	0.3300	< 0.020
Lufubu	0.9044	< 0.001

Appendix 5.8: Table of population specific ΦST values

	ΦST
Lufubu	0.71137
Congo-Kwango	0.68087
Upper Zambezi	0.70913
Okavango	0.70424
Middle Zambezi-Lower Zambezi-Shire	0.70642
Coastal	0.70726

Appendix 6: Reagent Recipes

Appendix 6.1: Recipe for Salt-Tris-EDTA

Ingredients:

- 5 mL 5 M Sodium chloride
- 5 mL 1 M Tris (pH 8.0)
- 1 mL 0.5 M EDTA (pH 8.0)

Bring the volume to 500 mL with ddH_2O

Appendix 6.2: Recipe for TBE Buffer

Ingredients:

- 108 g Tris
- 55 g Boric acid
- 40 mL EDTA (pH 8.5; 0.5 M)

Bring the volume to 1L with ddH_2O