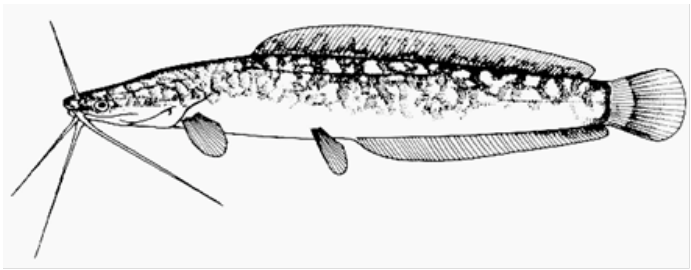
Faculty of Sciences

DEPARTMENT OF BIOLOGY

Unit of Research in Organismal Biology

Evolutionary genetics of the catfish *Clarias gariepinus* (Burchell, 1822) in the Congo Basin



A dissertation submitted by

Auguste CHOCHA MANDA

In partial fulfillment of the requirements

for the degree of PhD in Biological Sciences

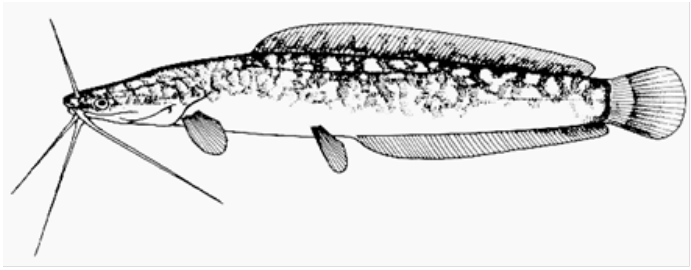
2010

Faculty of Sciences

DEPARTMENT OF BIOLOGY

Unit of Research in Organismal Biology

Evolutionary genetics of the catfish *Clarias gariepinus* (Burchell, 1822) in the Congo Basin



A dissertation submitted by

Auguste CHOCHA MANDA

In partial fulfillment of the requirements

for the degree of PhD in Biological Sciences

2010

Members of jury:

Prof. Jean-Claude Micha (Promoter, FUNDP)

Prof. Filip Volckaert (Co-promoter, K.U.Leuven)

Prof. Faustin Kang'Mate (Co-promoter, UNILU)

Prof. Patrick Kestemont (FUNDP)

Prof. Karine Van Doninck (FUNDP)

Dr. Maarten Larmuseau (K.U.Leuven)

Couverture: *Clarias gariepinus* (Burchell, 1822)

‘Eurêka’

Archimède de Syracuse

Remerciements

BTC BELGIAN DEVELOPMENT AGENCY

et

PRODEPAAK/BTCCTB

Au terme de cette recherche, je tiens à exprimer ma gratitude à tous ceux qui ont participé de près ou de loin à la réalisation de ce travail, que ce soit par une aide sur le terrain, au laboratoire, dans le traitement des données ou tout simplement pour les conseils et encouragements.

Je tiens à remercier tout particulièrement le Professeur Jean- Claude Micha, Promoteur de cette thèse, pour m'avoir accueilli il y a quelques années dans son laboratoire et pour la confiance qu'il m'a témoigné tout au long de cette recherche. Qu'il trouve ici l'expression de ma sincère reconnaissance.

Mes plus vifs remerciements s'adressent au Professeur Filip Volckaert, Co-promoteur de cette thèse, pour l'accueil et la formation dont j'ai pu bénéficier dans le laboratoire de Systématique et Diversité animale de la K.U. Leuven.

Un grand merci au Professeur Faustin Khang'Mate, Co-promoteur de cette thèse pour son soutien et son encadrement durant le séjour local au Congo.

Mes remerciements s'adressent aussi aux membres du Jury, notamment le Professeur Patrick Kestemont, Professeur Karine Van Doninck et Docteur Maarten Larmuseau, dont les remarques ont été d'un grand intérêt pour l'amélioration de ce travail.

Je pense que les mots ne suffiront pas pour remercier Bart Hellemans, sa disponibilité et son dévouement à me communiquer les bases de la biologie moléculaire ainsi que son aide dans la réalisation des travaux de laboratoires ont été d'un grand intérêt. Qu'il trouve ici ma sincère reconnaissance.

Que Lise-Marie Pigneur 'Maman Mapasa' et son mari Nicolas trouvent ici ma profonde reconnaissance pour tous les services bénéficiés que je ne saurais énumérer vu leur grand nombre. Grand merci.

Je pense aussi à mon ami Victor Pwema, à qui je souhaite du courage et de la persévérance sur ce long chemin.

Je désire également remercier les professeurs du D.E.S en Aquaculture. Je pense particulièrement aux Professeurs J-P. Descy, P. Devos, C. Mélard et J-P. Thomé. J- C. Philipart et P. Kestemont.

Mes remerciements s'adressent aux anciens et nouveaux membres de l'Unité de Recherche en Biologie des Organismes, des FUNDP, pour leur compagnie et leur

collaboration. Je pense à Pascal Isumbisho, Pascal Masilya, Gersande, L. Viroux, Robert Mandiki, Y. Mine, I. Imoro Toko, Y. Abou, Alou.

Mes remerciements s'adressent aux membres du Laboratoire de Systématique et Diversité animale de la K.U. Leuven pour leur collaboration, je pense particulièrement à Greg, Tine, Joost, Jo-Ann, Els, Eveline, Pascal, Nellie, Dirk, Frederik, Maartern Van Steenberghe, Anton, Sara, Alexander, Alex, Alessia, Jeroen Van Houdt, Maarten Vanhove, Prof Joss Snoeks.

Mes remerciements s'adressent au Professeur F. Nyongombe qui a conduit mes premiers pas dans ce monde merveilleux d'aquaculture.

Un grand merci au Doyen de la Faculté des Sciences Agronomiques de l'Unilu, le Prof Michel Ngongo pour ses encouragements.

Que mes amis de longue date trouvent ici ma profonde gratitude pour leur soutien, il s'agit de : Jules Nkulu, Emmery Kasongo, Adelin Meli, Vieux Jeff Mukalay, Willy Kaselekela, A. Banza.

Que les différents membres de la Faculté des Sciences Agronomiques de l'Unilu trouvent dans ces mots mes vifs remerciements, il s'agit de: Pr A. Ilaka, I. Tshibangu, F. Kampemba, L. Nyembo, D. Kasalwe, B. Mujinya, F. Munyemba, M. Ngoy Schutscha, B. Katemo, P. Muteba, P Bassay, L Mbaya, S. Kalala, L. Mbiya, L. Kapongo, M. Mpundu, A. Mukoj, O. Eanga, D. Mwamba, A. Lubobo. J. Lwamba, D. Fwana, E. Ngoma, Sr C. Nsiami.

Que les membres du Projet de Développement de la Pêche et Aquaculture Artisanale du Katanga trouvent ici notre profonde gratitude. Nous citons, Jean-Pierre Marquet et sa famille, Emile Mukuba, Mathurin Tshishi, Jean Tshilay, Da Nyota, Da Françoise, Pasteur Kashala, Dieudo Sangwa, Kabey.

Un grand merci à A. Kafumbu, F. Bilo, A. Katim, Papa Mushota, Papa Banza, J-P Nday Brido, Directeur Musubao pour leur soutien et aide sur terrain.

Que le Professeur Mutambwe qui ne cessait de m'encourager trouve ici ma profonde reconnaissance.

Que mes encadreurs à la bourse mixte BTCCTB, Monsieur C. Misigaro, J-C Kakudji et A. Mowa, trouvent ici ma profonde reconnaissance pour tous les services.

Il me semble que je dois aussi remercier mes proches. Ainsi, je pense que ma grande Sœur Mado et mon grand Frère Timothée, seront très contents de l'aboutissement de ce long parcourt, fruit de leurs sacrifices. Que mes grands frères Jean et Felix ainsi que mes sœurs Frida, Albertine et Mimi se sentent soulager après plusieurs années de sacrifices.

Que mon épouse A. Katempa ainsi que mes enfants les 'Manda' trouvent dans ce travail un couronnement et le fruit des sacrifices consentis.

Que mon grand frère Pascal Kalunga et sa femme Pascaline ainsi que toute sa famille, trouvent dans ces quelques mots mes sincères remerciements.

Que mon Ami Maarten Vanhove ainsi que ses parents trouvent dans ces quelques mots les sentiments d'amitiés.

Un grand merci à la famille Volckaert, pour l'accueil dont j'ai bénéficié lors de mon passage à la maison.

Que mon ami Peter Mangwala, ainsi que sa famille trouve ici l'expression de ma gratitude.

Enfin ce travail, je le dédie à mon défunt père, à ma défunte mère et à mon défunt frère François.

Merci à tous

Bedankt iedereen

List of Abbreviations

AFLP	Amplified fragement length polymorphism
AIC	Akaike information criterion
BI	Bayesian inference
bp	Base pair
BTCCTB	Belgian development agency
cyt *b*	Cytochrome *b*
$d\mu^2$	R_{ST} linked pairwise genetic distances according to Goldstein *et al.* (1995)
D_{CE}	F_{ST} linked pairwise genetic distance according to Cavalli-Sforza & Edwards (1967)
DNA	Deoxyribonucleic acid
D-test	Tajima's (1989) neutrality test
EST	Expressed sequence tags
F_{IS}	Inbreeding coefficient
F_{ST}	Fixation index
Fs-test	Fu's (1997) neutrality test
GIS	Geogrphic information system
GSM	Generalized Stepwise Mutation model
h_d	Haplotype diversity
H_e	Expected heterozygosity
$H_{e.n.b}$	Unbiased expected heterozygosity
H_o	Observed heterozygosity
HWE	Hardy-Weinberg equilibrium
IAM	Infinite allele model
IBD	Isolation by distance
Indels	Insertions/deletions
kYA	kilo Years Ago
K	Numbers of clusters calculated in Structure
KAM	K Allele Mutation model
LD	Linkage disequilibrium
LGM	Last glacial maximum
Ln P (D)	Estimated log probabilities

LRT	Likelihood-ratio test
MCMC	Markov chain Monte carlo
MDSA	Classical multidimentional scaling analyses
ML	Maximum likelihood
mtDNA	Mitochondrial DNA
MYA	Million years ago
N	numbers of samples
N_H	Number of haplotypes
NJ	Neigbour-joining
p	P-value
PCR	Polymerase chain reaction
PRODEPAAK	Projet de Développement de la Pêche et de l' Aquaculture Artisanale au Katanga
P_S	Number of polymorphic sites
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction fragment length polymorphism
R_{ST}	Analogue of F_{ST} with taking account of the stepwise mutation model
SMM	Stepwise mutation model
SNPs	Single nucleotide polymorphism
SSRs	Simple sequence repeats
T3P	Tamura 3-Parameter
θ	Estimator of F_{ST}
π	Nucleotide diversity
ϱ	Estimator of R_{ST}

Table of Contents

List of figures

List of tables

Evolutionary genetics of the catfish *Clarias gariepinus* (Burchell, 1822) in the Congo Basin

Abstract

The Congo Basin features among the oldest, largest and biologically richest on earth after the Amazon and before the Mekong. Sadly enough it is the least known. It offers opportunities to study evolution given its treasure of global biodiversity and high aquaculture potential (large hydrographic network and suitable climate). It is within this framework that phylogeographical research has been initiated to understand the origin of the ichthyological fauna and to identify potential strains for the development of aquaculture. The model chosen is the catfish *Clarias gariepinus* (Burchell, 1822), a species with great aquaculture potential. This fish lives in the Congo River from source to mouth, as well over a large part of the African continent. The results show the presence of four clades and several groups which represent a large genetic differentiation. Evolution is ancestral in the south and north of the Congo Basin (relict population) and represents a Pleistocene colonization in the Central Basin. Even in a contemporary context, evolution seems to follow these two models. Two hypotheses have been retained to support the causes for the accumulation of diversity: the hypothesis of long-distance dispersal in the Central Basin and the refuge hypothesis in the peripheral zones of the Central Basin. Moreover, the four major clades have been identified to match with four major ichthyological provinces of Africa: Nilo-Sudan, Zambezi, East Coast and Congo. This confirms an ancient connection between the various basins. Hence, the evolution of *C. gariepinus* in the Congo basin corresponds with that of the fishes and mammals with a high dispersal capacity, which is found in several African basins. Ideally the evolution of the endemic fauna of the Congo basin should be studied in order to compare it with the higher mentioned high dispersal group. It would facilitate our understanding of the evolution of the fish fauna of this large African basin for sustainable management.

Evolution génétique du poisson chat *Clarias gariepinus* (Burchell, 1822) dans le bassin du Congo

Résumé

Le bassin du Congo est l'un des plus vieux, vastes et biologiquement riches au monde après celui de l'Amazone et avant celui du Mekong. Malheureusement ce bassin est le moins connu. En tant que réserve de la biodiversité mondiale et ayant un potentiel aquacole élevé (grand réseau hydrographique et bon climat), ce bassin offre des opportunités d'études d'évolution. C'est dans ce cadre que des études phylo-géographiques ont été initiées afin de connaître l'origine de la faune ichthyologique. Le modèle retenu était celui du poisson chat *Clarias gariepinus* (Burchell, 1822). Ce poisson est signalé dans le fleuve Congo de la source à l'embouchure, ainsi que dans la plus grande partie du continent africain. Les résultats obtenus ont révélé l'existence de quatre clades (groupements historiques de populations) rencontrés dans le bassin du Congo et plusieurs groupes présentant une grande différenciation génétique. Une évolution ancestrale (relique) caractérise le sud et le nord du bassin congolais; par contre dans la Cuvette Centrale on observe une colonisation récente. Même dans un contexte contemporain, l'évolution semble suivre les deux modèles. Deux hypothèses ont été retenues pour soutenir les causes de l'accumulation de la diversité. Il s'agit de l'hypothèse sur la longue distance de dispersion dans la Cuvette Centrale et celle des refuges dans les régions périphériques de la Cuvette Centrale. En plus, les quatre clades principaux ont été identifiés comme couvrant les quatre principales provinces ichthyologiques d'Afrique: Nilo-Soudanique, Zambèze, Côte Est et Congo. Ceci confirme l'ancienne connexion entre les différents bassins. Ainsi, l'évolution du *C. gariepinus* dans le bassin du Congo correspond à celle des poissons et mammifères à grande dispersion rencontrés dans plusieurs bassins africains. L'idéal serait d'étudier aussi l'évolution de la faune endémique du bassin congolais afin de la comparer avec celle appartenant au premier groupe (grande dispersion). Cla facilitera la compré-hension de l'évolution de la faune ichtyologique de ce grand bassin d'Afrique pour un but de gestion durable.

Chapter 1

General introduction

1.1. The diversity and evolution of Africa's freshwater systems

The inland waters of Africa and Madagascar are extremely diverse both in the form and life they support. The largest rivers system have the most diverse fauna, but faunal composition does not exhibit dramatic changes across the intertropical area (Lévèque, 1997). At the heart of Africa, the Guinean-Congolian forest rivers are some of the biologically richest waters on the continent, with species adapted to rapids, swamp forest, large and small rivers, and large lakes (Kamdem Toham *et al.,* 2003).

The biogeography of the African freshwater fish remains poorly known in comparison with Europe and North America, which may be explained by the considerable taxonomic and ecological complexity of the ichthtyofauna (which is not yet well-known), the limited expertise and research effort but also by the poor fossil record (Lévêque, 1997). The African ichthyofauna contains over 3,200 species belonging to 94 families. This is impressive as an estimated 11,000 species inhabit the freshwater habitats of the world (Lévêque & Paugy, 2006). African fish fauna, as we observe today, evolved over time. Historically, the faunal barriers between the hydrographical systems of Africa during the early Miocene, some 25 million years ago, were less distinct than today. Climate and geology during the late Miocene and Pleistocene shaped the local habitats (Maley, 2001) and fauna (deMenocal, 2004). Climatic fluctuations with a succession of dry and wet phases contributed to the delineation of the distinct ichthyological provinces, and consequently the geographical distribution of fish species (Roberts, 1975). The peculiar case of the radiation of cichlid fishes in the African Great Lakes bears convincing testimony. Fragmentation in small lakes during the dry period of the Pleistocene led to allopatric evolution.

Based on the affinities of the fish fauna between different regions, the inland waters of Africa and Madagascar are subdivided into ichthyofaunal provinces. Thus, Roberts (1975) and Lévêque (1997) recognized the following ichthyofaunal provinces in

Africa and Madagascar: the Maghreb, Nilo-Sudan, Upper Guinea, Lower Guinea, Congo, the Quanza, Zambezi, East Coast, Southern and Malagasy province (Fig 1.1). The Maghreb province has an extremely poor fish fauna, only five families have been reported (Lévèque, 1997). Its freshwater fauna shares a strong affinity with Europe's Mediterranean ecoregions. The origin of the fish fauna is Palearctic with some tropical elements (Thieme *et al.,* 2005). Here lives the only native Salmonidae species (*Salmo trutta*) in Africa (Roberts, 1975).

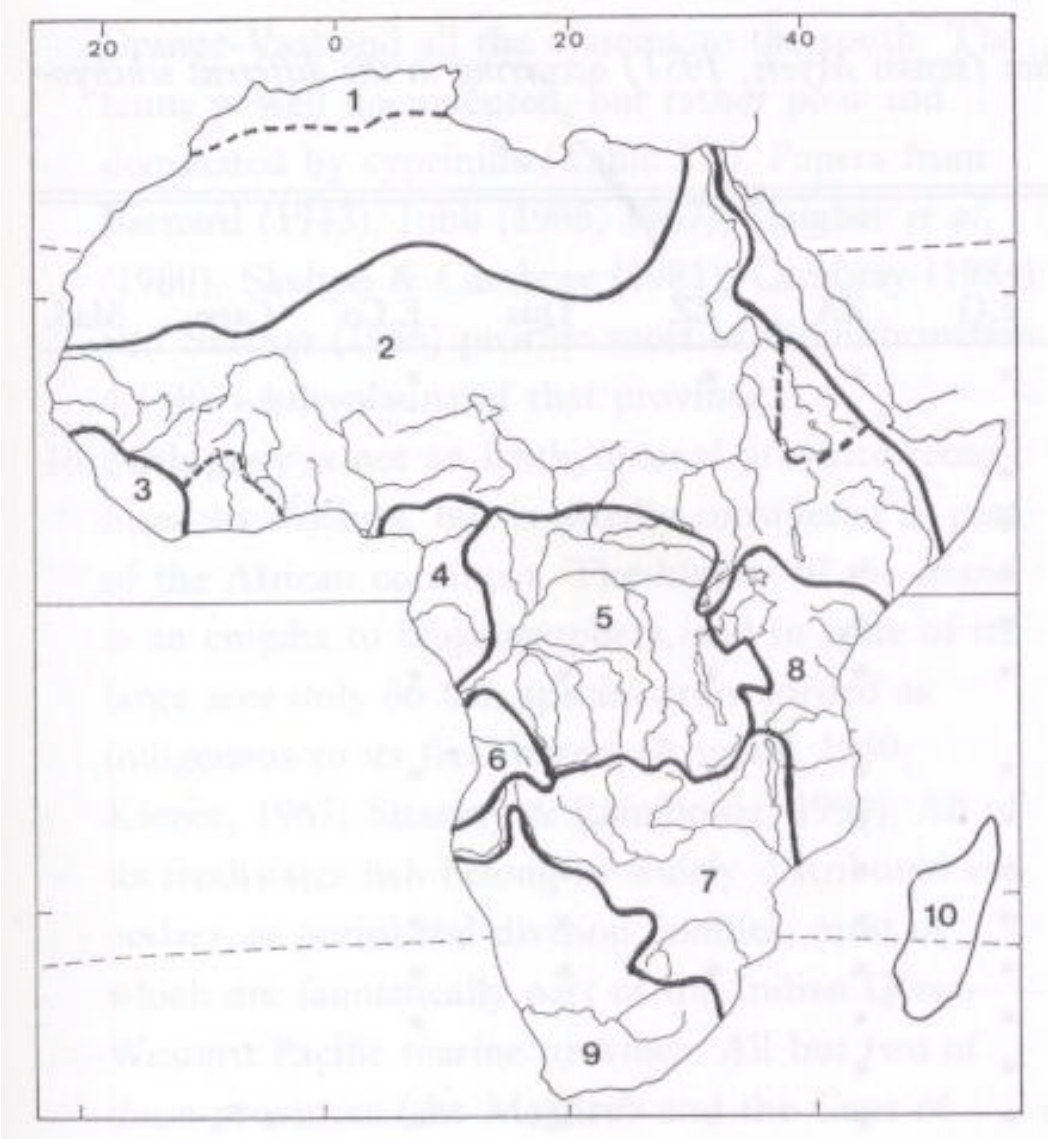


Figure 1.1. Main ichthyological provinces in Africa according to Roberts (1975) and Lévèque (1997): (1) Maghreb, (2) Nilo-Sudan, (3) Upper Guinea, (4) Lower Guinea, (5) Congo, (6) Quanza, (7) Zambezi, (8) East Coast, (9) Southern and (10) Malagasy provinces.

The Nilo-Sudan province includes the major drainage basins of the Nile, Chad, Niger, Volta and Senegal (Lévèque, 1997; Thieme *et al*., 2005; Lévêque & Paugy, 2006). The Upper Guinea includes coastal rivers from south of the Kogon River in Guinea to Liberia, and exhibits faunistic affinities with the lower Guinea province and the Congo (Lévèque, 1997, Roberts, 1975). The Lower Guinea covers the coastal rivers from Cameroon to the mouth of the Congo (which is not included) (Lévèque, 1997). The Quanza province covers the Angolan coastal drainages (Lévèque, 1997). Data on this region is very poor. The Zambezi province, includes the river basins of the Cunene, Ovambo, Okavango, Zambezi and Limpopo, as well as Lake Malawi (Lévèque, 1997). The eastern basin, which includes the middle and lower Zambezi, Limpopo and the Shire River, connected with the Congo until the formation of lakes in the Rift Valley (Lévèque & Paugy, 2006). The East Coast province covers the coastal drainages from the Juba in the north to the Zambezi in the south (Lévèque, 1997). The Southern province includes the basins of the Orange-Vaal and all the systems to the south. The fauna is poor and dominated by cyprinids (Lévèque, 1997). All freshwater fish of the Malagasy province belong to widely distributed secondary or peripheral divisions families. The Congo province includes, according to Roberts (1975) and Lévêque & Paugy (2006), Lake Kivu and Tanganyika. Connections between the Congo basin and adjacent basins have been mentioned in the literature (Lévèque & Paugy, 2006). It includes exchanges with the Zambezi (Bell cross, 1966; Skelton, 1994; Katongo *et al*., 2005) and Chad basin (Lévèque, 1997).

1.2. The Congo basin

1.2.1. Characteristics of the Congo basin

The surface of the Congo basin is estimated at 4,000,000 km^2 (Lévèque & Paugy, 2006), an area almost equal to the Indian subcontinent. This international basin has waters flowing in the states of Angola, Burundi, Cameroon, Central African Republic, Republic of Congo, Democratic Republic of Congo, Rwanda, Tanzania and Zambia (Lévêque, 1997; Thieme *et al*., 2005). The Congo Basin has the highest species richness of any river system on the African continent and is second after the

Amazon Basin at a global level. Its enormous surface and high diversity of habitats have facilitated the evolution of a highly diverse freshwater fauna. Geological events that have increased the number of barriers to faunal exchange have also provided opportunities for species evolution. Unfortunately, information on the biodiversity of the Congo Basin is very limited. Nevertheless, the entire Congo Basin has been delineated in nineteen freshwater ecoregions by Thieme *et al*. (2005), namely Bangwelu-Mweru, Lake Tanganyika, Lake Kivu, Thysville Caves, Malagarasi-Moyowosi, Upper Lualaba, Kasai, Uele, Mai Ndombe, Tumba, Central Basin, Lower Congo, Malebo Pool, Sangha, Sudanic Congo, Upper Congo, Albertine highlands, Lower Congo Rapids and Upper Congo Rapids.

The Congo River is divided into three major sections (Roberts & Stewart, 1976; Banister, 1986) (Fig 1.2.):

- The Upper Congo, also called Lualaba, from its source to Kisangani (Wagenia waterfalls). The river is characterized by waterfalls and rapids, such as the 'Porte de l'Enfer' near Kongolo.
- The central Congo basin, *Cuvette Centrale* or middle Congo, which goes from Kisangani to Kinshasa. The river is wide and crosses the vast central plains. This portion receives the largest tributaries, such as the Ubangi and Kasai rivers.
- The Lower Congo, form West of Kinshasa (Pool Malebo) to the mouth. This part is characterized by rapids and falls, and an estuary upon its entry in the Atlantic Ocean.

The D.R. Congo, which is situated in this basin, covers the largest part of tropical forest of Central Africa (some one million km^2) (Fig 1.2). Here reside more than 10,000 plant species, 3,000 of which are known to be endemic. It is also home to at least 409 species of mammals, 1086 species of birds, and localized groups of culturally and genetically unique human forest inhabitants (Anon, 1991). The south of the D.R. Congo houses the high species richness and endemism across the Katanga-Chambeshi region (Cotterill, 2005). Especially the Katanga region appears a hotspot with its endemisms of mammals, such as the antelope *Kobus anselli* Cotteril, 2005 (Cotterill,

2005) and fishes such as *Parakneria lufirae* Poll, 1965 and the cichlid *Oreochromis salinicola* Poll, 1948 (Thys van den Audenaerde, 1964; Poll, 1976).

The freshwater biodiversity threat has been reported to be low in the Congo basin (Vörösmarty *et al*., 2010) although some threats do exist, such as pollution from mining, forest exploitation, fishing and hunting (Thieme *et al*., 2005).

1.2.2. Fish diversity

The number of fishes known from the Congo basin continues to increases. Poll & Gosse (1963) reported more than 408 primary freshwater fish arranged in twenty-four families (Thieme *et al.*, 2005). Lèveque & Paugy, (2006) reported more than 1000 freshwater fishes (787 species for Congo river and tributaries and 291 species for Lake Tanganyika) arranged in thirty-eight families. River systems or ecoregions at the periphery of the Congo Basin harbour distinct faunas (Poll, 1963, 1976; Magis, 1961). For example three cuvette centrale families are absent in the Lualaba (Phractolaemidae, Notopteridae and Pantondidae). Cichlids are uncommon in the main rivers, and only the large fluviatile genus *Thylochromis* is well represented. The specialized cichlid *Orthochromis torrenticola* Thys van den Audenaerde, 1963 lives in rapids and close to waterfalls in the Lufira river (Banister, 1986). The small endemic *Barbus* species *Barbus kamalondonesins* Poll, 1938 inhabits only the Upper Lualaba. The endemic African family Kneriidae counts two genera in the Congo Basin: *Kneria* and *Parakneria*. Fishes of this family are usually regarded denizens of high altitude streams only, for example two endemic species *K. wittei* Poll, 1944 and *K. katangae* Poll, 1976 occur up to 1800 m while *P. lufira* Poll, 1965 lives at a lower altitude of 600-800 m (Banister, 1986; Bailey, 1986; Poll, 1976).

The faunisticaly relatively well-known regions (Figure 1.3.) are limited to the rapids downstream of Kinshasa (Roberts & Stewart, 1976), Lefini river (Zamba & Verven, 2008), Pool Malebo (Poll, 1959), Yangambi (Poll and Goss, 1963), Lake Tumba (Matthes, 1964), Upemba region (Poll, 1976), the Luapula-Mweru system (de Kimpe, 1964), Inkisi river (Wamini Lunkayilakio *et al.,* 2010), Lake Tanganyika (Poll, 1986) and Lake Kivu (Snoeks *et al.,* 1997).

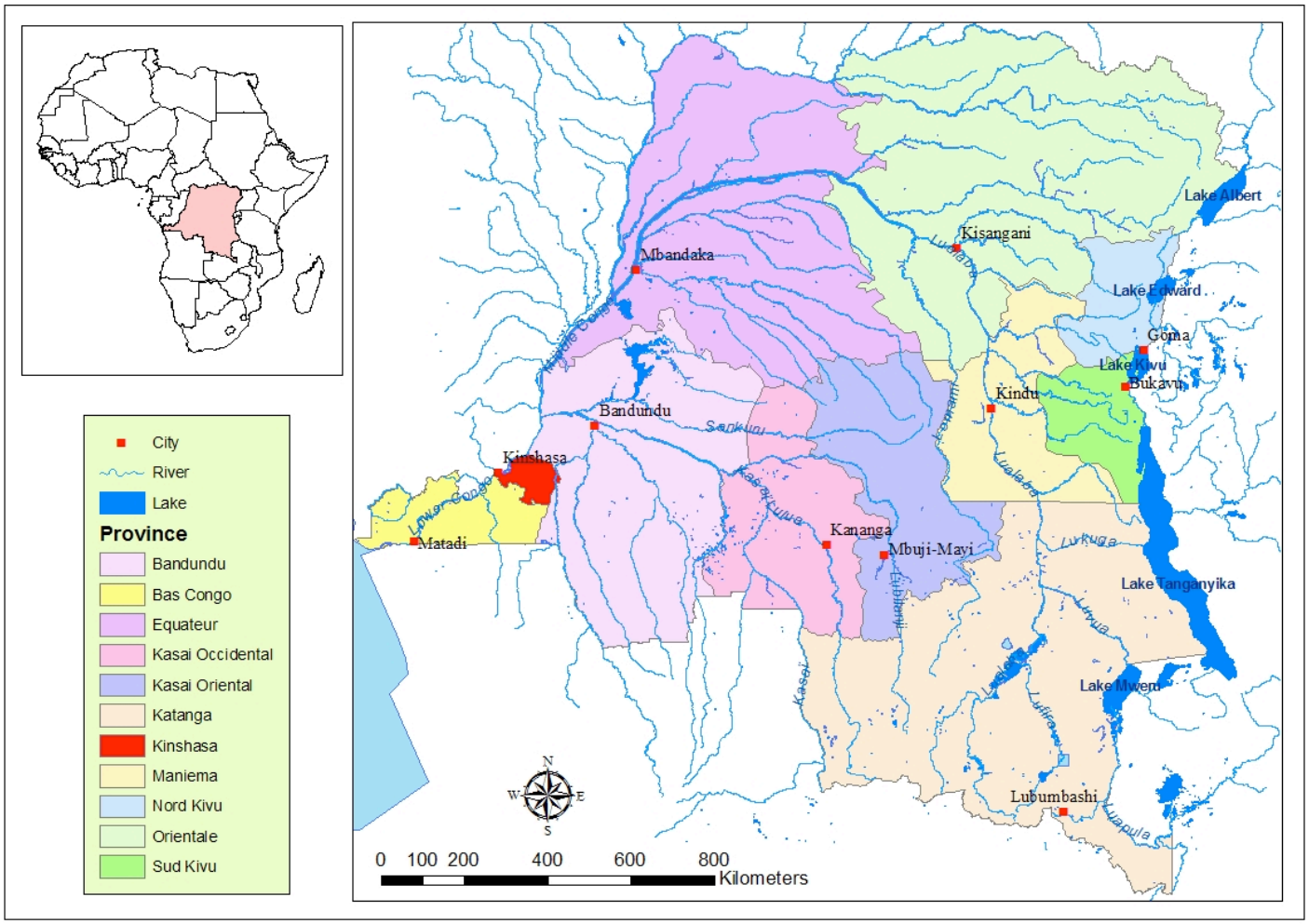


Figure 1.2 Geo-political map of the Democratic Republic of Congo

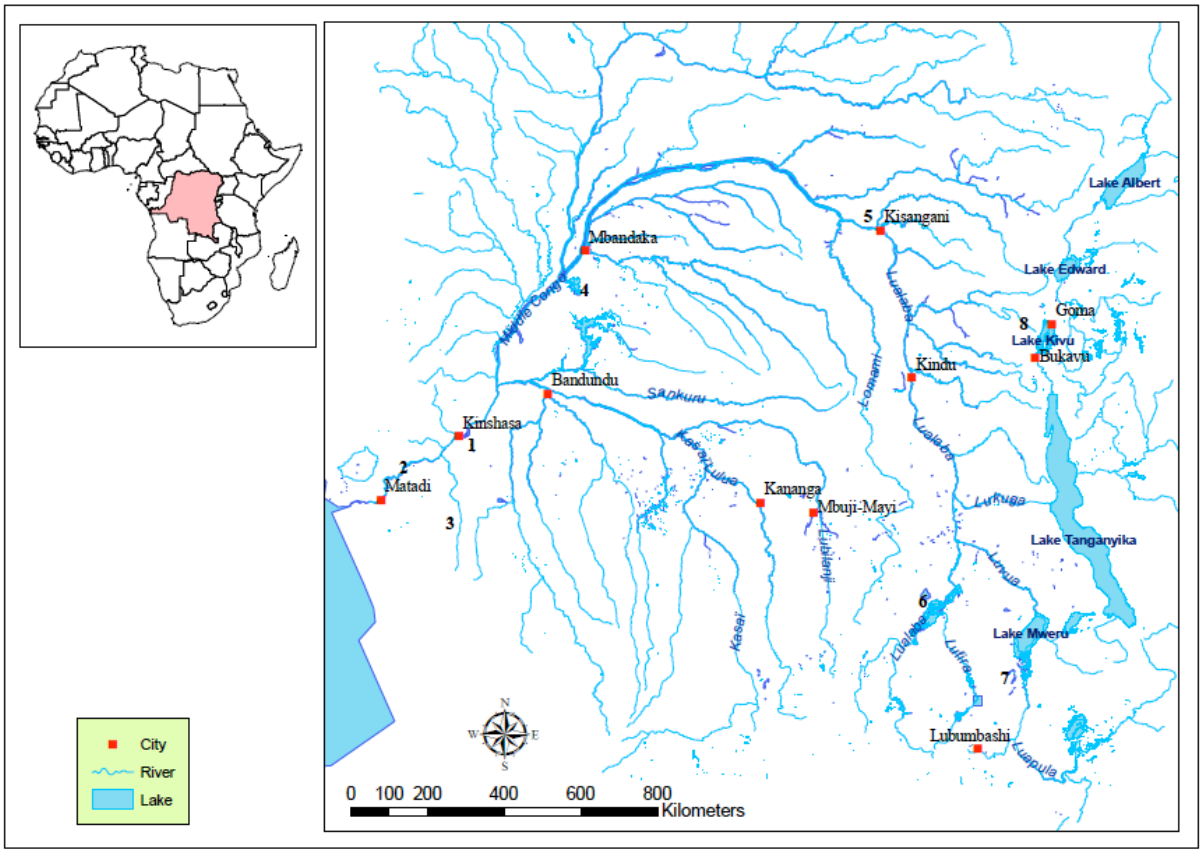


Figure 1.3 The Congo basin with indication of faunisticaly relatively well-known sites and rivers: (1) Pool Malebo, (2) the rapids below Kinshasa, (3) Inkisi river, (4) Lake Tumba (5) Yangambi, (6) Upemba region, (7) Luapula-Mweru system and (8) Lake Kivu.

1.2.3. Exploitation

The Democratic Republic of Congo has large water surfaces, which represent an exploitation potential of 707,000 tons of fish annually. The current production of fish is estimated at 250,000 tons by capture fisheries and 2690 tons of *Oreochromis niloticus* from aquaculture (FAO, 2007). Production has not followed the rhythm of demographic growth; hence t. The production deficit is covered by import.

1.2.4. Speciation hypotheses

The Congo Basin represents an extraordinary site for evolutionary studies. It harbours a high level of species richness and endemism. Several hypotheses were proposed to study origin of species richness in other large tropical rivers (Amazon and Mekong rivers) (Fig 1.4). The Museum, Paleogeography, River (fragmentation of biota) and Hydrogeological (dispersal barriers) hypotheses (Hubert & Renno, 2006; Hubert *et al.,* 2007; Adamson *et al.,* 2010) offer a perspective on how faunas have accumulated and evolved.

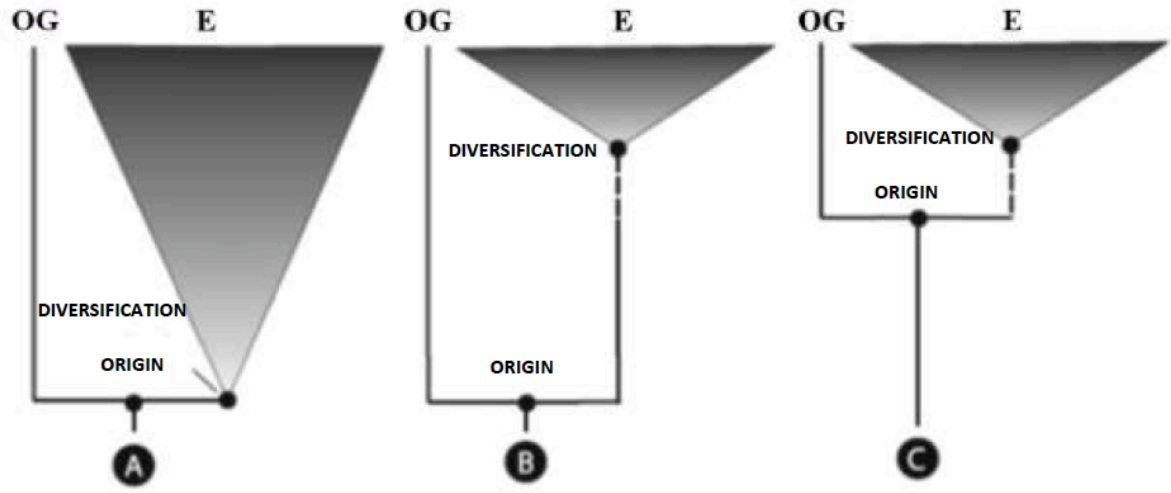


Figure 1.4. Alternative phylogenetic patterns over time and their possible interpretation (OG: outgroup and E: endemic). (A) Museum model, (B) Mountain refugia, Island refugia and Long-distance dispersal model (C) Long-distance dispersal model (adapted from Murriene, 2009).

The Museum hypothesis is based on the principle that the modern species originated by allopatric differentiation in a stable region (e.g. mountain forests) during marine

highstands. Later on they accumulated by dispersal in the lowlands, which act as 'museums' (Hubert & Renno, 2006; Murrienne, 2009). It has been used to explain the origin of neotropical fish in the Amazon basin (Hubert *et al*., 2007).

The *Paleogeography hypothesis* is based on the principle that the modern species originating by allopatric differentiation follow the changing landscape in response to the tectonic fluctuations (Haffer, 2008). The lakes of the African rift valley are an example (Lévèque, 1997; Giddelo *et al*., 2002; Lévèque & Paugy, 2006).

The River hypothesis is based on the principle that speciation is due to the barrier effect of rivers (e.g. water falls) which caused isolation for a long time (Haffer, 2008, Katongo *et al.,* 2005, 2007).

The Hydrological hypothesis is based on the principle of the speciation due to fragmentation of the river during dry periods (Ardnt *et al*., 2003, Hubert *et al*., 2007).

The *Long-distance dispersal hypothesis (large dispersal model)* suggests that the present species originate from neighboring region (Murienne, 2009).

We will test these hypotheses in the Congo Basin with the phylogeographic method. However few organisms have the advantage that their large-scale geographical history allows to put phylogeographical processes in a continental perspective. Such a case is the catfish (*Clarias gariepinus* Burchell, 1822) with its Pan-African distribution and common occurrence in the Congo Basin (Teugels, 1986). It represents a good model to test these hypotheses.

1.2.5. The research model: the catfish Clarias gariepinus)

Taxonomy and systematic classification

The catfish *Clarias gariepinus* (Burchell, 1822) belongs to the family of Clariidae (Fig.1.5). This family is one of the 31 families belonging to the order of the Siluriformes (Teugels, 2003). It is distinguished from others by the absence of the spines at the end of the dorsal fin.

Phylum: Chordata
Subphylum: Vertebrata
Superclass: Gnathostomata
Superclass: Osteichthyes
Class: Actinopterygii
Subclass: Neopterygii
Infraclass: Teleostei
Superorder: Acanthopterygii
Order: Siluriformes
Family: Clariidae
Genus: *Clarias* Scopoli, 1777
Species: *Clarias gariepinus* (Burchell, 1822)

Figure 1.5. Taxonomy of *Clarias gariepinus* (Burchell, 1822)

The dorsal and anal fins are very long (Fig 1.6.); the fish has four pair of barbels and the presence of an accessory breathing organ enables it to breath air when very active or under very dry conditions. They live in the muddy substrates of ponds and occasionally gulp air through the mouth (Mbega & Teugels, 2003).

The genus *Clarias* (Scopoli, 1777) is found on the African and Asian continents (Teugels, 2003). *C. gariepinus* belongs to the group of species that have relatively long barbels which, in addition to their gustative function, play the role of tactile receptors. The group shows a close relationship between *C. gariepinus* and *C. anguillaris* (Jansen *et al.,* 2006). Also, *Bathyclarias sp*., the result of a recent speciation event in Lake Malawi (Agnèse & Teugels, 2001b) has a paraphyletic status. Contrary to other *Clarias* species, *C. gariepinus* has a large number of gill rakers varying from 24 to 110, the number increasing with the size of the fish; these gill rakers are long, slender and closely set. A high correlation between this number and the standard length characterizes the species (Teugels, 1986; Benech et *al*., 1993).

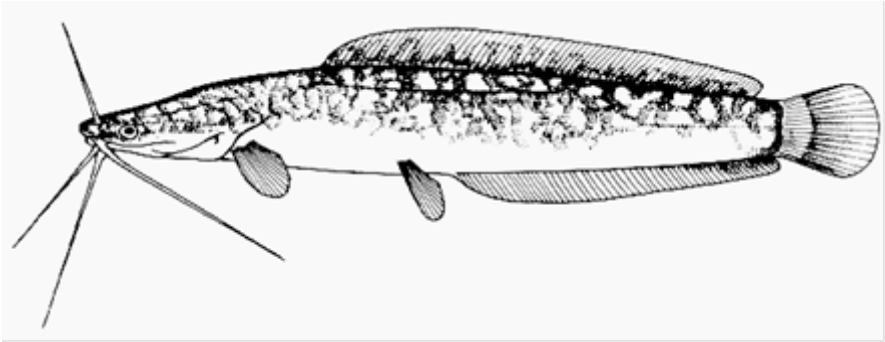


Figure 1.6**.** *Clarias gariepinus* (Burchell, 1822).

Distribution

Representatives of the family of the Clariidae may be found (Fig1.7.) anywhere in Africa, in the Middle-East, and in certain parts of Asia (Vitule *et al.*, 2006, Jansen *et al*., 2006). The original distribution of *C. gariepinus* in Africa stretches from the Nile and the Middle East to West Africa and from Algeria to South Africa (De Graaf & Janssen, 1996; Teugels, 1986) where it lives in streams, lakes, rivers of which some are seasonally dry (Janssen, 1985). Fish survive by the respiration accessory organs that allow respiration from air oxygen (Viveen *et al.*, 1985; De Graaf & Janssen, 1996).

During these last 20 years, the species has been introduced in Europe, Asia and South America (Ducarme & Micha, 2003; Poompuang & Na-Nakorn, 2004) for aquaculture production. In Europe it is used for intensive aquaculture, whereas in Asia it is used for hybridization with local species. In fact, the hybrid of *C. gariepinus* and *C. macrocephalus* is popular in the Asian aquaculture community because of its rapid growth rate and resistance to diseases (Lal *et al.*, 2003; Sahoo *et al.*, 2003; Senanan et *al*., 2004).

Ecology and Biology

Clarias gariepinus lives mainly in quiet waters, lakes and pools but may also occurs in fast flowing rivers and rapids (Teugels, 1986; Skelton & Teugels, 1992).

In nature, *C. gariepinus* matures after 1 to 3 years and the length varies between 150 and 800 mm (Hecht *et al.*, 1988). In pond aquaculture it reaches sexual maturity between 7 and 10 months (Viveen *et al.,* 1985) and, with some exceptions, doesn't reproduce spontaneously. In this case reproduction is induced artificially (Janssen, 1985).

First sexual maturity occurs when females are between 400-450 mm and males between 350-400 mm. *C. gariepinus* is oviparous and spawning takes place in flooded area. Fish make a lateral migration towards the inundated plains to breed and return to the river or lake soon afterwards while the juveniles remain in the inundated area. Eggs are greenish. Incubation takes little time (about 33 h at 25°C) (Janssen, 1985).

Juveniles return to the lake or river when they are between 15 and 25 mm long (Teugels, 1986).

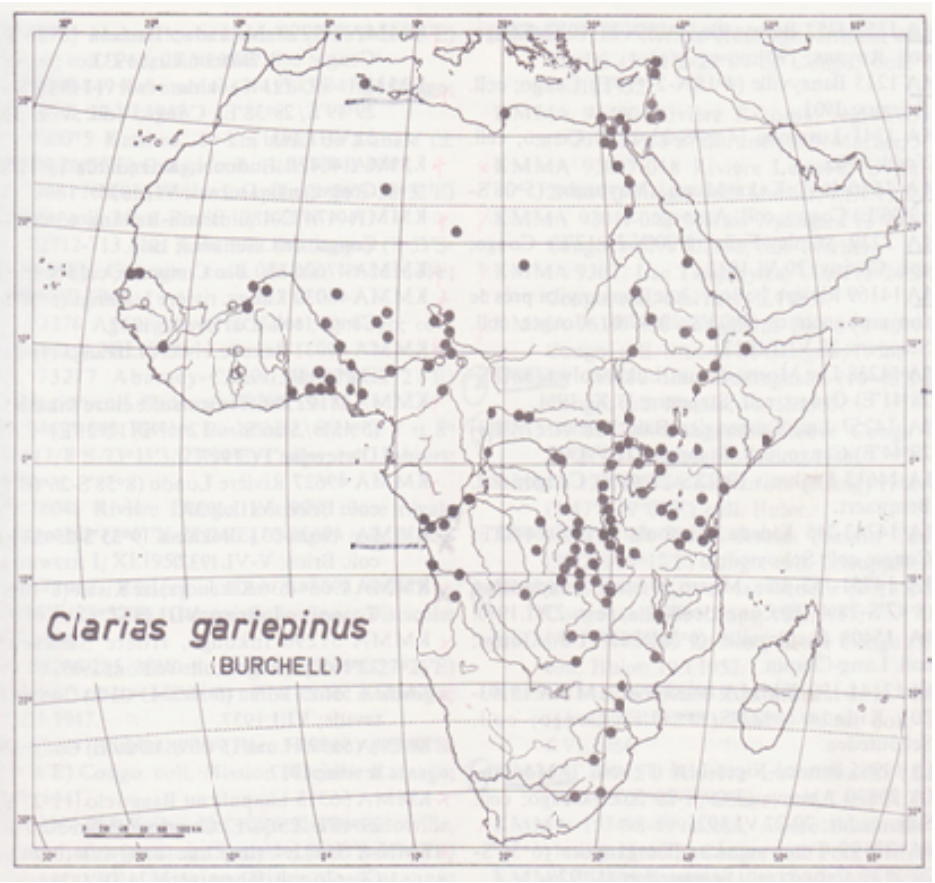


Figure 1.7. Map of the disctribution of the catfish *Clarias gariepinus* (Teugels, 1986)

C. gariepinus is considered as an opportunistic omnivore. The food regime of *C. gariepinus* changes according to the size of the fish (Yalçin *et al*., 2001). The food of *C. gariepinus* is composed of fishes, mollusks, insects, other plants, and zooplankton (Dadebo**,** 2000). Many organisms prey on *C. gariepinus*, such as the fish *Hydrocynus vittatus* (Alestidae) (Mhlanga**,** 2003) and the insect *Belostomata sp* (Notonectidae) (Adeyemo *et al*., 1997). In aquaculture catfish show cannibalistic behavior under suboptimal living conditions.

Aquaculture

Given its growth potential, catfish is an established fish for African aquaculture. The large production outside Africa and the availability of extensive to intensive strategies

show that production can be achieved on the African continent. Many aquaculture projects have been initiated in Africa to develop the breeding of catfish (Central African Republic: Micha, 1973; Janssen, 1985, Republic of Congo: Janssen, 1985, Republic of South Africa: Hetch *et al.*, 1988: Nigeria: Nwadukwe, 1995). It is highly appreciated across the continent and fetches higher prices on the market of big towns.
Catfish has been a cultured intensively and extensively outside Africa in Europe (Belgium, Italy, Germany, The Netherlands, Poland and U.K), Asia (Bangladesh, Indonesia, Peoples Republic of China, Philippines, Thailand and Vietnam), South America (Brazil) and has been abandoned in the Middle East (Israel) (Huisman & Richter, 1987; Verreth & Eding, 1993; Hecht *et al.*, 1996; Więcaszek *et al.*, 2010). African production stagnates (Ducarme & Micha, 2003) due to the poor knowledge of methodology for the production of fry at the level of the African farmer, the lack of knowledge on the diversity and performance of the strains of *C. gariepinus*, and also regional access to appropriate feed. The largest producing countries in the world are Brazil (America), Cameroon, Ghana, Kenya, Mali, Nigeria, Republic of South Africa and Uganda (Africa), Syria (Middle East), Peoples Republic of China (Asia). The global production is 33,924 tons for a value of 100.6 million USD (FAO, 2010).

Fish farming started in the D.R. Congo in 1940 and was abandoned after independence in 1960. Pond production reached then 5 tons per hectare per year. Tilapia (*Oreochromis machrochir* and *Tilapia rendalli*) were the first species domesticated (De Bont, 1950). Nowadays, interest is growing again but the choice of the adequate fish species or strain remains a problem. The catfish *C. gariepinus* is the second most important freshwater fish after *Tilapia* in the region. It has proven merits for African aquaculture, especially for its fast growth, high disease resistance, high density and low feed conversion (Haylor, 1993).

1.3. Phylogeography as a tool to study biodiversity

Too few studies have focussed on fish evolution in the Congo Basin, although it represents one of the most important organisms with more than 1000 species (Lévèque & Paugy, 2006). The numerous phylogeographic studies on mammals (Comstock *et al.*, 2002; Quérouil *et al.*, 2003; De Vivo & Carmignotto, 2004; Hewitt, 2004; Johnson

et al., 2007; Tosi, 2008; Moodley & Bruford, 2007) and the few studies on fish (Joyce *et al*., 2005; Katongo *et al*., 2005; Koblmuller *et al*., 2008; Sturmbauer *et al*., 2005) have made it possible to understand the impact of the Congo river on the evolution of species during the Pleistocene. Several scenarios have been suggested to explain the large biodiversity. Poll (1963, 1976) suggested the hypothesis that de river Lualaba (upstream Congo) ran to the Nile in the past. He based this on the presence of fish species which were shared between the Congo and Nile basins. The conclusion of Poll (1963, 1976) after the biogeographical study of the fishes of the south of Congo was that the affluents found in the upstream part of the Congo river (Luapula, Lufira, Lualaba) did not belong to this basin in the past. Indeed the affinities between the fauna of the Zambezi river and the Congo basin are striking. The region represent the delineation between the Congo and the Zambezi, as has been confirmed in a study on the local Cichlidae (Katongo, 2005, 2007). The evolution of *Clarias gariepinus*, which is considered native to the Congo basin and is found from source to mouth, may provide an answer to the different hypotheses which have been published on the origin of the ichthyofauna of the Congo Basin.

Phylogenetic and phylogeographical studies of closely related species based on molecular markers such as mitochondrial DNA (mtDNA) are expected to provide insights into the history of a region. When DNA substitution rates are known, the timing of speciation events may be estimated from genetic distances and compared to known geological events. Likewise, when species distribution ranges are known, the phylogenetic biogeography may provide a comprehensive picture of the evolutionary history of a region (Hubert *et al,* 2007). In complement other molecular marker such as allozymes, microsatellites, Amplified Fragment Length Polymorphisms, Restriction Fragment Length Polymorphisms and Single-Nucleotide Polymorphisms generate valuable information (Liu & Cordes, 2004). They are discussed below.

1.3.1. Molecular genetic markers

A wide range of markers are used in population genetics (Table 1.1). They have been used in order to establish the degree of heterozygosity within a population to identify parentage, to evaluate the genetic distance between natural populations or domesticated strains to determine the link with related species, to detect loci coding for quantitative and qualitative characters, to map the genome and many other applications.

Allozymes

Allozymes are allelic variants of proteins produced by a single (or more) gene locus; they are of interest as markers because polymorphism exists and because they represent protein products of genes. Hence they are considered type I markers. Disadvantages associated with allozymes include heterozygote deficiencies due to null alleles and the amount and quality of tissue samples required (Liu & Cordes, 2004). In addition, some changes in DNA sequence are masked at the protein level, reducing the level of detectable variation. Many studies in the past revealed low levels of genetic variation (Liu & Cordes, 2004). Van der Walt *et al*. (1992), Van der Bank *et al*. (1992), Van der Walt *et al*. (1993) and Na-Nakorn *et al*. (2002) have used allozymes for the study of diversity among populations of *C. gariepinus.* Their results reveal average heterozygosities ranging between 0.0033 and 0.075 in population from The Netherlands, South Africa and Thailand (Table 1.2).

Table 1.1. Types of DNA markers, their characteristics, and potential applications (Liu & Cordes, 2004)

Marker type	Acronym or alias	Requires prior molecular information?	Mode of Inheritance	Type	Locus under Investigation	Likely allele Numbers	Polymorphism or power	Major applications
Allozyme		Yes	Mendelian, Codominant	Type I	Single	2 – 6	Low	Linkage mapping, population studies
Mitochondrial DNA	mtDNA	No	Maternal Inheritance	-		Multiple haplotypes		Maternal lineage
Restriction fragment length polymorphism	RFLP	Yes	Mendelian, Codominant	Type I or Type II	Single	2	Low	Linkage mapping
Random amplified polymorphic DNA	RAPD, AP-PCR	No	Mendelian, Dominant	Type II	Multiple	2	Intermediate	Fingerprinting for population studies, hybrid identification
Amplified fragment length polymorphism	AFLP	No	Mendelian, Codominant	Type II	Multiple	2	High	Linkage mapping, population studies
Microsatellites	SSR	Yes	Mendelian, Codominant	Mostly Type II	Single	Multiple	High	Linkage mapping, population studies, paternity analysis

Table 1.1. (continued)

Marker type	Acronym or alias	Requires prior Molecular information?	Mode of Inheritance	Type	Locus under Investigation	Likely allele Numbers	Polymorphism or power	Major applications
Expressed sequence tags	EST	Yes	Mendelian, Codominant	Type I	Single	2	Low	Linkage mapping, physical mapping, comparative mapping
Single nucleotide polymorphism	SNP	Yes	Mendelian, Codominant	Type I or Type II	Single	2, but up to 4	High	Linkage mapping, population studies?
Insertions/deletions	Indels	Yes	Mendelian, Codominant	Type I or Type II	Single	2	Low	Linkage mapping

Table 1.2. Origin of samples, Average Heterozygosity and author

Origin of samples	Average Heterozygosity	Author
Thailand	0.055	Na-Nakorn *et al*. (2002)
The Netherlands	0.0033-0.0759	Van der Walt *et al.* (1993)
South Africa	0.0033/0.047	Van der Bank *et al.* (1992)

Microsatellite DNA

Microsatellites or simple sequence repeats (SSRs) represent codominant molecular genetic markers, which are ubiquitously distributed within genomes (Chiastiakov *et al*., 2006). Their polymorphism is based on the variation in the number of repeats of a simple DNA sequence (2 to 6 nucleotides long). Microsatellites have gradually replaced allozymes as the preferred tool of detecting fine-scale structuring and for testing hypotheses regarding the biological significance of such as particular environmental, geographical, or demographic factors (Hardy *et al*., 2003; Larsson *et al*., 2007; Ruzzante *et al,*. 2006).

Several problems are associated with generating and interpreting microsatellite data, thus which statistics to use (mutation models, null alleles, stuttering, large allele, sampling size required).

Mutation model: The choice of good model of mutation is important for statistical analysis for will reflect a probably genetic structure. Comparison of microsatellite alleles can thus provide two kinds information: allele identity/ nonidentity and allele size differences. Two groups of models of mutation have been proposed to describe variation at microsatellite loci. First, the model group based on allele identity difference (identity in state) includes: the Infinite Allele Mutation model (IAM), the K Allele Mutation model (KAM). Second, the model group based on alleles size difference includes: the Stepwise Mutation Model (SMM) and the Generalized Stepwise Mutation model (GSM) (Table 1.3.) (Hardy *et al.,* 2003). The IAM predicts that mutation will lead to only new allelic states and may involve any number of repeat units; consequently, this model does not allow for homoplasy. In contrast, the SMM predicts that each mutation creates a novel allele either by adding or deleting a single repeated unit of the microsatellite; consequently, this model has a memory of allele

size and does allow for homoplasy. Most statistics that describe genetic differentiation from genetic markers (F-statistics) rely solely on allele identity information. This information is often used to infer phylogenetic relationships or obtain indirect estimates of gene flow. Two parameter (F_{ST} and R_{ST}) can been used for to describing the degree of genetic differentiation among populations. F_{ST} is defined as the correlation of allelic states and R_{ST} is an analog of F_{ST} based on allele size difference.

Table 1.3. Mutation models (Hardy *et al.,* 2003)

Models	**Effect of a mutation event**
IAM: infinite-allele model	New allele (never observed previously) created
KAM: *K-allele model*	Mutation toward one of *K possible* allelic states (excluding the original state)
SMM: stepwise mutation model	Allele size increased or decreased by just 1 unit
GSM: generalized stepwise model	Allele size modified by x unit, x being a random variabel following any distribution of finite variance

Balloux & Goudet (2002) suggested that F_{ST} (IAM) is efficient in the case of high levels of gene flow and R_{ST} (SMM) is better reflects population differentiation under low gene flow.

Null alleles: Null alleles are non-amplified alleles that, when segregating with another allele, resulting in false homozygotes (Pemberton *et al*., 1995; Van Oosterhout *et al*. 2004). The effects of null alleles appear to be particularly clear when populations differ in null allele frequencies (Van Oosterhout *et al*., 2004). *Stuttering:* this phenomenon is related, when the microsatellite alleles are observed as a series of bands, and not a single discrete band (scoring of stutter peaks**)**. Large allele dropout refers to the increased likelihood of PCR failures for larger products and is often related to poor template quality.

Sampling size and number of loci required: The results of microsatellite variability are sensitive to the number of loci scored, the number of populations sampled, and the number of individuals typed in each sample (Evanno *et al.,* 2005). The sampling size

required depend on the species and the locus investigate. A minimum sample size of 50 individuals per population should be considered for loci showing between five to ten alleles (O'Connell & Wright, 1997; Evanno *et al*, 2005).

Galbusera *et al*. (1996) developed seven specific microsatellite primers for *Clarias gariepinus* to perform paternity tests and characterize wild and domesticated populations. Senanan *et al*. (2004) developed three additional microsatellites from *C. gariepinus* and analyzed the genetic impacts of hybrid catfish farming *(Clarias macrocephalus x C. gariepinus*) on native catfish populations in central Thailand. The number of alleles per locus was comprised between 4 to 8 and heterozygosities ranged from 0.34 to 0.89. Microsatellites isolated from other species such as *C. batrachus* (Volckaert & Hellemans, 1999; Yue *et al*., 2003; Islam *et al.,* 2007) and *C. macrocephalus* (Na-Nakorn *et al.,* 1999; have been used for the characterization of populations of *C. gariepinus.* Microsatellites isolated from *C. batrachus* showed a lower number of alleles than those from *C. gariepinus* (Table 1.3) despite the high number of loci (18) used.

Table 1.4. Heterozygosities and Allelic diversity at different microsatellite loci isolated from *C. gariepinus* and *C. batrachus* used in the evaluation of genetic variation within and among populations of *C. gariepinus* collected from central, northern and southern Thailand (H_o = Observed heterozygosity*;* H_e = Expected *heterozygosity;* $H_{e.n.b}$ = Unbiased expected heterozygosity**).**

Origin of microsatellite	**Number of loci**	**H_o**	**H_e**	**$H_{e.n.b}$**	**Number of alleles**	**Author**
C. gariepinus	7	-	-		5.00-14.00	Galbusera *et al.* (1996), Volckaert and Hellemans (1999)
	7	0.13- 0.78	0.18- 0.79	0.18-0.79	2.14-10.71	Galbusera (1997)
	6	0.50- 0.69	0.67-0.75		4.67-12.17	Wachirachaikarn *et al.* (2009)
C. batrachus	18	0.17-1.00	0.17- 0.85		1.00-6.00	Yue *et al.* (2003)

Mitochondrial DNA

Generally sequence divergence accumulates more rapidly in mitochondrial than in nuclear DNA (Brown, 1985). This has been attributed to a faster mutation rate in mtDNA that may result from a lack of repair mechanisms during replication (Wilson *et al*., 1985) and a smaller effective population size due to the strictly maternal inheritance of the haploid mitochondrial genome (Birky *et al*., 1989). Analyses of mtDNA marker have been used extensively to investigate the identity of taxa (DNA barcording) in a stock structure in a range of fishes and to estimate effective population size. These studies have provided a better understanding of the origin of fishes species in the Amazon basin (Renno *et al.,* 2006).

A disadvantage of mtDNA is that this marker shows maternal inheritance; the phylogenies and population structures derived from mtDNA data may not reflect those of the nuclear genome due to gender-biased migration. In addition, mtDNA markers are subject to the same problems that exist for other DNA-based markers, such as back mutation (sites that have already undergone substitution are returned to their original). Many researches have used this marker to study the history of the modern geographical distribution of fishes such as the *C. gariepinus* in Africa (Table 1.2.). Arndt *et al*. (2003) used the D-Loop and related the fragmentation of *C. gariepinus* in Northen Africa to wet/dry periods during the Pleistocene. Agnèse & Teugels (2001b) showed with the *cytochrome b* gene that the *Bathyclaris sp* of Lake Malawi and *C. gariepinus* of the Luapula and Lake Malawi share a common ancestor. Giddelo *et al*. (2002) used the ND5/ND6 gene to establish the structure of *C. gariepinus* in eastern Africa. The authors show the presence of tree distinct groups (East, North and South-central).

Table 1.5. Structure of populations of *Clarias gariepinus* based on analyses of a fragment of mtDNA.

Origin of samples	Control region or fragment of mtDNA	Structure of population	Author
Middle East (Turkey, Syria, Israel), North Africa (Egypt), Western Africa (Senegal, Mali and Niger)	*D-Loop*	Fragmentation of population in northern Africa	Arndt *et al*. (2003)
Lake Malawi and Luapula River (Zambia)	*Cytochrome b*	*Bathyclarias sp* (they share a common ancestor) from Lake Malawi and *C. gariepinus* from Luapula River and Lake Malawi are form a monophyletic group	Agnèse & Teugels (2001b)
Lake Victoria (Kenya)	*Cytochrome b*	All haplotypes of tributaries rivers of Lake Victoria founded in this Lake	Mwita & Nkwengulila (2008).
Eastern , North and the South-central of Africa	*ND5/6*	Presence of tree distinct genetic groups (East, North and South-central)	Gidello *et al.* (2002)
India and Thailand	*ND 5/6*	Phylogenetic tree reaveled and *C. macrocephalus*. three distinct clusters *C. batrachus*; *C. gariepinus* (India and Thailand);	Mohindra *et al*. (2007)

- *ND5/6*: (gene provides instructions for making a protein called NADH dehydrogenase 5/6)
- *D*-Loop: mtDNA control region

Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphism (AFLP) is a DNA fingerprinting technique that does not require prior sequence knowledge or genetic information of the species (Palti, 2009). Weaknesses are the dominant inheritance imposing enormous difficulties on the transfer of information and the additional laborious steps that are required in developing single-locus markers from informative polymorphic AFLP fragments (Felip *et al*., 2005). Poompuang & Na-Nakorn (2004) used this technique to obtain a preliminary genetic map of walking catfish (*Clarias macrocephalus*). The level of polymorphism was relatively low, because few polymorphic bands (4.3 bands per primer pair) were observed.

Restriction Fragment Length Polymorphism (RFLP)

Restriction Fragment Length Polymorphism (RFLP) is based on resolving length differences of DNA fragments from restriction enzyme digestion using gel electrophoresis (Palti, 2009). A major advantage of RFLP is that they are codominant markers. However, they are much less polymorphic than microsatellites and compared to SNPs they are much less abundant in the genome and less adaptable for automated genotyping (Palti, 2009; Liu & Cordes, 2004). Giddelo *et al*. (2002) detected the genetic structure of *C. gariepinus* at the ND5 and ND6 loci (gene provides instructions for making a protein called NADH dehydrogenase 5/6) of mitochondrial DNA using RFLP–PCR. These results show the presence of three clades in eastern of Africa, the East clade, the North clade and the South-central clade. The nucleotide diversity is comprised between 0.000 and 0.0660.

Single-Nucleotide Polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs) describe polymorphisms caused by point mutations that give rise to different alleles containing alternative bases at a given nucleotide position within a locus (Liu & Cordes, 2004; Morin *et al*., 2004; Brumfield *et al*., 2003). Today, next-generation sequencing allows the discovery of large numbers of single nucleotide polymorphisms (SNPs) in species where little genomic information was previously available (Renaut *et al*., 2010). This marker that proved to

be very useful in studies of population genetic and conservation (Shawn *et al.,* 2010, Joanne *et al.,* 2009, Phillip *et al.,* 2009). Sauvage *et al.* (2007) with SNPs obtain a very high level of DNA polymorphism in oysters. The comparison of Single Nucleotide Polymorphism with Short Tandem Repeat and Allozymes on genetic differentiation of population of Chinook Salmon showed for overall divergence (F_{ST}), a higher value for SNPs than for other markers (Smith *et al.,* 2007). Narum *et al*. (2008) used the SNPs and microsatellite markers for the differentiating salmon population at broad and fine geographical scales. The results showed that both types of markers are likely to be useful in population genetics studies and that, in some cases, a combination of SNPs and microsatellites may be the most effective suite of loci. Technological advances make SNPs genotyping still a challenging endeavor and requires specialized equipment (Liu & Cordes, 2004).

In conclusion, mtDNA has a relatively fast rate of nucleotide divergence, well suited for examining events over the last few million years. For more recent events (last 10,000 years) other markers, such as microsatellites, SNPs, are more suitable (Hardy *et al*., 2003, Hewitt, 2004, Renaut *et al*., 2010).

1.3.2. Population demography

Phylogeography takes into account the demographic history because of the impact on population structure and genealogy. The demographic history of population includes population growth, population reduction or population constancy. It can be evaluated by:

(1) Genetic diversity: Haplotype diversity is a measure of the uniqueness of a particular haplotype in a given population (Nei and Tajima, 1981), Nucleotide diversity is a concept which is used to measure the degree of polymorphism within a population (Nei, 1987).

(2) Other demographic tests: Deviation from neutral expectations may be evidence for past demographic events such as population growth or the presence of population sub-structuring within a sample. In this case we used the demographic test such as Tajima's *D,* Fu's F_S and R_2. Tajima's *D* (Tajima, 1989) that compares the number of segregating sites to nucleotide diversity in a sample (Class I) was used to test for deviation from neutrality due to selection, population bottleneck, or admixture (Rand, 1996). Fu's F_S (Fu, 1997), that compares θ estimated from nucleotide diversity with the expected number of haplotypes under Ewens' (1972) distribution given the sample size (Class II) was used to detect past fluctuation in population size, and R_2 (Ramos-Onsins &Rozas 2002), that uses information from the mismatch distribution of pair-wise differences (Class III) was also employed to examine demographic events.

1.3.3. Population genetics

Population genetics is the study of allele frequency distribution and change under the influence of the four main evolutionary processes: natural selection, genetic drift, mutation and migration (Hartl & Clark, 1997). It also takes into account the factors of population subdivision and population structure. It attempts to explain such phenomena as adaptation and speciation. The population genetics is based on a testable null model or hypothesis which is called the Hardy-Weinberg Equilibrium (HWE). Both allele and genotype frequencies are constant in the absence of evolution. This principle is useful for testing the sources of evolution (migration, genetic drift, mutation and natural selection).

1.4. Research questions

This study covers the historical evolution of the fauna of the Congo Basin. The key question asks whether the evolution of a fish species with a large geographical distribution may help to interpret the evolution of the local fauna and the impact of past hydrogeological and climatological modifications. The genotypes required for the analysis came from tissue samples of correctly identified *Clarias gariepinus* collected across its range in the Congo basin. In a first part, (**Chapter 2**) for the first time ever the question was addressed whether the Congo basin harboured a single or several

clades of the common fish *C. gariepinus*. Phylogenetic relationships and phylogeographic patterns based on the *cytochrome b* locus were evaluated. Divergence times were estimated between the clades. Once the historical pattern was established, the question arose whether contemporary patterns diverged from the historical patterns (**Chapter 3**). A study of the genetic structure based on highly variable microsatellite markers was performed on eight populations from the Congo Basin. Genetic variation and differentiation among and between populations were analysed and compared with the historic patterns. As *C. gariepinus* lives across the whole African continent, the analysis of the historical pattern as detected in the Congo Basin was expanded to the full continent (**Chapter 4**). A larger dataset of *cytochrome b* sequences was evaluated phylogenetically and phylogeographically, and compared with the well established ichthyological provinces. In the final chapter the importance and the implication for genetics conservation are discussed.

Chapter 2

Phylogeography of the catfish *Clarias gariepinus*: a window on the origin and evolution of the Congolese ichthyofauna

A. Chocha Manda, M. H. D. Larmuseau, M. P. M Vanhove, B. Hellemans, F. Khang'Mate, J.-C. Micha and F. A. M. Volckaert

Submitted to the Journal of Fish Biology

Abstract

The distributional patterns and genomes of the fish fauna of the Congo basin should carry the traces of successive periods of wet and moist climate regimes during the Pleistocene. Although circumstantial evidence is considerable, local proof has remained scant. Here the phylogeography of the widespread catfish *Clarias gariepinus* (Clariidae, Teleostei) is analyzed in the Congo basin and the adjacent upland regions based on a fragment of the mitochondrial cytochrome *b* locus in 118 individuals. Twenty four unique sequences clustered in two clades: the central basin proper (Congo clade) and the southern Lulua-Kando (Kasaï) clade. Network analysis revealed the presence of several clusters in the Congo clade (Congo-Lualaba, Luapula-Mweru, Upper Lufira and Middle Lufira), and additional clusters in the north (Ubangi) and east (Tanganyika-Rusizi and Kivu) of the Congo basin. The upper Luapula and Lake Mweru region, which has been the scene of numerous river captures, have a high haplotype diversity. Most remarkable is the low genetic diversity of the Congo basin proper (Cuvette Centrale), with a single haplotype group occupying the region from the upper Lualaba (Kamalondo depression) downstream up to Kinshasa. The genetic pattern revealed an early Pleistocene differentiation of *C. gariepinus* in the east, north and south, and a late Pleistocene colonization of the Congo basin following its drainage some 400 kYA. Our results present a complementary view on the aquatic ecoregions of the Congo basin.

2.1. Introduction

Processes such as continental drift, climate change, geological activity, hydrography, biota and stochastic dispersal events drive the distribution patterns of continental organisms. While the Congo Basin has the highest species richness of any river system on the African continent (Thieme *et al*., 2005), only second to the Amazon Basin, the contribution and dynamics of these patterns remain poorly understood. Moreover, the African forest and savannah harbour an assemblage of species of varying age, hence the usefulness of studying phylogenies and phylogeographies (Plana, 2004).

The Central Congo Basin (Cuvette Centrale), formerly a continental basin, was invaded by the sea several times during the Mesozoic (250-67MYA). During the Pleistocene glaciations (2.6 MYA to 12 kYA) with their arid periods, forests became fragmented and organisms retracted into refuges (Plana, 2004). Although dry phases during the Pleistocene altered the vegetation of the Congo Basin, the river system has remained stable (Hughes & Hughes, 1992). During periods of lower temperatures (sub)montane floras descended into the lowlands. About 400 kYA, a coastal river, which is now called the Congo River, returned to tap the huge Lake Congo (Beadle, 1981; Stankiewicz & de Wit, 2006). The environmental stability of the area over a long period, the wide range of habitats, and the long isolation from other bioregions are thought to have favored the evolution of rich biota, including many endemic species (Poll, 1976; Beadle, 1981; Banister, 1986; Lévêque, 1997; Thieme *et al.*, 2005).

River systems at the periphery of the Congo Basin harbour distinct faunas. The fauna of the Luapula-Mweru system, which has historically been part of the Zambezi system during the early Tertiary and is known as the Zambian-Congo branches of the Congo system (Skelton, 1994; Thieme *et al*., 2005), shows the least affinity with the Congo Basin (Poll, 1963, 1976) (Fig. 1.1). The faunal affinities of the Kamalondo depression, also known as Upemba depression (Upper Lualaba river) are more pronounced to the Central Basin than either to the Luapula system or to the upper tributaries of the Kasai system. Relatively more rheophilic species inhabit the Katanga plateau in the south of

Congo basin compared to the alluvial plain of the Kamalondo depression (Upemba), Luapula and upper tributaries of the Kasai-Lulua (Poll, 1976). The numerous falls and rapids along the course of the Kasai River and its tributaries may explain the high level of endemism (Beadle, 1981; Thieme *et al.*, 2005). Some upstream tributaries of the Kasai system belong to the Zambezian headwater ecoregion as a consequence of a historic connection (Poll, 1976; Thieme *et al*, 2005). The Great Lakes located in the east, Kivu and Tanganyika, are the results respectively of a relatively recent volcanic eruption (Virunga) during the Pleistocene (15 kYA) and tectonic past during the Miocene (12-9 MYA) (Lévêque, 1997). Particularities of Lake Kivu are its depauperate fish fauna comprising just 28 species (Snoeks *et al*., 1997), from which the haplochromine cichlid fishes of Lake Victoria are derived (Verheyen *et al*., 2003). An ancient Lake Kivu connection with Lake Victoria has been implicated (Beadle, 1981). High habitat diversity and hydrogeographic barriers have shaped the ichthyofauna north of the Congo Basin (Ubangi drainage), with its affinity to the Nilo-Sudanic system (Thieme *et al*., 2005).

The Congo Basin includes several freshwater ecoregions (Thieme *et al*., 2005). Its biota has been shaped by climate oscillations throughout the Quaternary, with periods of extreme climate variability during wetter episodes (Trauth *et al*., 2009). During the Last Glacial Maximum, the climate in the tropics was colder and drier, with an increase in desert and savannah surface and a decline of the rain forest coverage. The tropical forest was fragmented in a restricted number of refuges (Maley, 2001; Plana, 2004), while in the savannah ecotone the vegetation shifted from a C3 (trees and shrubs) to a C4 metabolism (tropical grasses) between 1.5 and 0.7 million years ago (MYA) (Ségalen *et al*., 2007). These events lead to changes in African faunal assemblages with expansion of dry tropical species in three main areas: West, East and South Africa (Hewitt, 2004; Moodley & Bruford, 2007). The genomes of the fauna of the Congo Basin bear the imprint of these Pleistocene dynamics. For example mammals such as chimpanzees and elephants inhabiting the tropical forest went through periods of population expansion and reduction, when retracting into refuges (Anthony *et al*., 2007; Johnson *et al.,* 2007). This should also be reflected in the freshwater fish fauna, but the few studies have focused on regions at the edge of the

basin. Climate driven water level changes of the Great Lakes have measurably influenced the population sizes of fishes, notably haplochromid cichlids (Verheyen *et al.,* 2003; Salzburger *et al*., 2005), while river dynamics have impacted dispersal (Giddelo *et al*., 2002; Katongo *et al*., 2005, 2007; Koblmüller *et al*., 2006).

Few organisms have the advantage that their wide-scale geographical history allows to put the phylogeographical processes in a continental perspective. Such an example is the catfish *Clarias gariepinus* Burchell (Clariidae; Teleostei) with its pan-African distribution, common occurrence in the Congo Basin (Teugels, 1986), generalist life-style and broad habitat preference (rivers, lakes, streams, rivers, marshes and floodplains from the coastal zone up to 1500 m above sea level). It has evolved several key adaptations to survive the dry season (suprabranchial organ, skin breathing, movement over land and short embryonic development time; Teugels, 1986; Skelton & Teugels, 1992). Previously the species was subdivided into three taxa, with *C. lazera* Valenciennes living in central and western Africa, *C. gariepinus* in eastern Africa and *C. mossambicus* Peters in southern Africa. Ozouf-Costaz *et al.* (1990) synonymized all three taxa, although anecdotal evidence points to morphological, ecological and physiological differences between the regions.

Generalities on the historical genetic diversity of *C. gariepinus* on a pan-African scale have appeared in the literature. All three major genetic clades (northern Africa, eastern Africa and South-Central Africa) have been spotted in the region of the Congo basin. The Ubangi population (Central African Republic) had been grouped with the North-African clade (Agnèse & Teugels 2001b), the Upper Congo samples from the Rusizi River (Burundi) and Luiche River (Tanzania) have been allocated to the eastern clade (Giddelo *et al*., 2002) and the West-Zambian Congo samples have been assigned to the South-Central clade (*sensu* Giddelo *et al*., 2002) by Agnèse & Teugels (2001b).

This study is the first to document specifically the clade structure of a common inhabitant of the Congo Basin. It addresses the historical population structure of *C. gariepinus* in the drainage basin of the Congo River. It is hypothesized that the phylogeography of a common fish is representative of the local fauna because of the shared hydrological and climatological dynamics.

2.2. Materials and Methods

2.2.1. Sampling and mtDNA sequencing

A total of 118 adult *C. gariepinus* individuals were collected at 24 sites in the Congo Basin, spread all over the provinces of Katanga, Orientale, Kinshasa and Kasai in the Democratic Republic of Congo (Fig. 1.2.) between September 2006 and January 2010 (Fig. 2.1, Table 2.1). Live fish were purchased from local fishermen, identified on site and barbell clips collected for preservation in pure ethanol (100%).

Genomic DNA was extracted from barbell clips using a NucleoSpin®96 Tissue Kit (Macherey-Nagel GmbH Co.KG) following the manufacturer's instructions. The primers used, L15267, 5'-AAT GAC TTG AAG AAC CAC CGT-3' and H15891, 5'-GTT TGA TCC CGT TTC GTG TA-3' (Briolay et *al*., 1998), amplified a fragment of 660 bp of the mitochondrial cytochrome *b* (cyt *b*). PCR reactions were performed in a volume of 25 µl containing 2.5 µl of 10xPCR buffer (Eurogentec), 1 µl 50 mM $MgCl_2$ (Eurogentec), 2.5 µl of 2 mM dNTPs (Amersham Pharmacia Biotech), 1 µl of each primer (20 µM), 0.1 µl of 5 U/µl Silverstar Taq polymerase (Eurogentec) and 16 µl of milli-Q H_2O. Cycling conditions were as follows: initial denaturation at 95°C for 3 min, 35 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 45 s, and a final extension step of 72° for 7 min. After purification from the agarose gel with the GFX Purification Kit (Amersham Pharmacia Biotech), sequencing was done using the same primers as above, with the Big Dye Terminator 3.1 kit (Applied Biosystems), applying a 1/8 dilution of the Big Dye Terminator sequencing protocol. Products were finally run on an ABI 3130 Genetic Analyser (Applied Biosystems). Sequence data were analyzed using the sequencing analysis software SeqScape v.2.5 (Applied Biosystems). Sequences were deposited in GenBank under accession numbers XXX-XXX *(numbers available after revision)*.

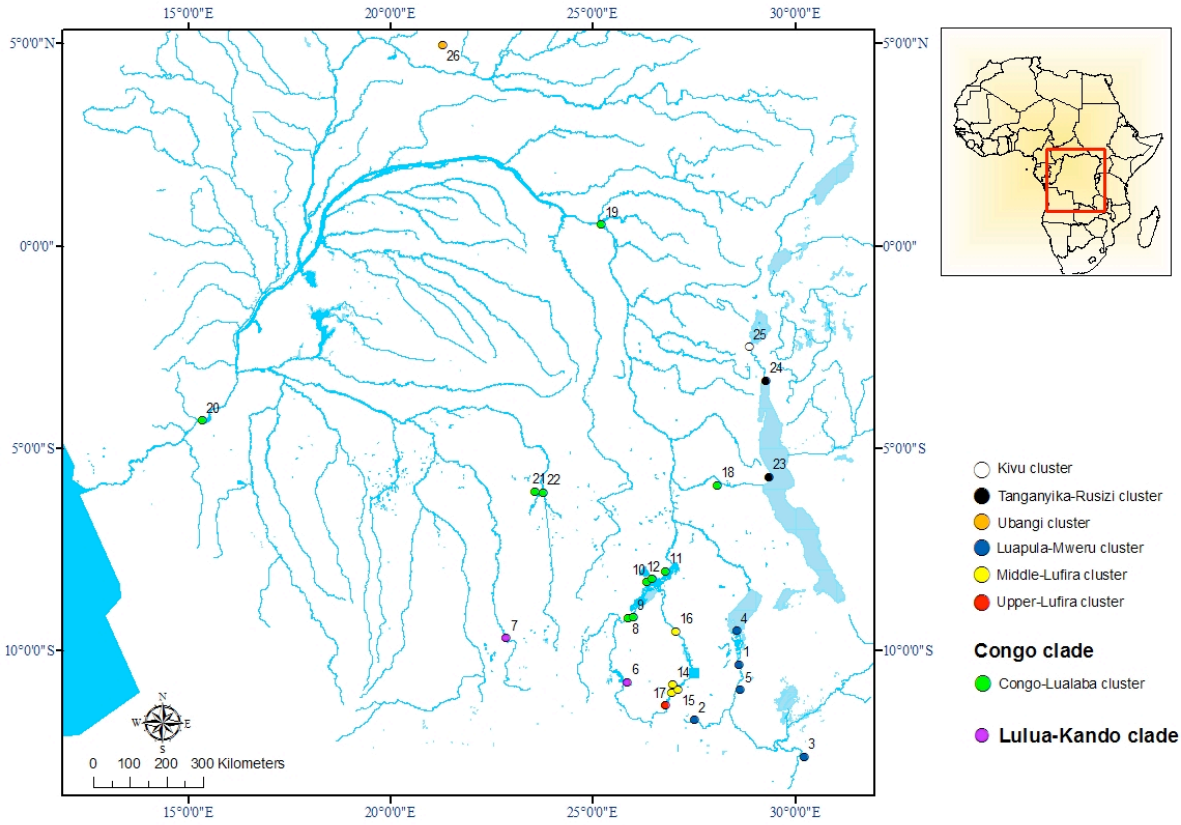


Figure 2.1. Geographical origin of the samples of *Clarias gariepinus*. Codes correspond to codes in Table 1. Colours refer to the clade or clusters designation.

Table 2.1. Code, geographical origin of the samples of *Clarias gariepinus* (river or lake, location and drainage basin, Provinces of D R. Congo, geographical coordinates), number of samples examined (N) and unique cytochrome *b* haplotypes. Locality numbers correspond to the numbers in Fig 2.1.

Site #	Code	River or lake	Location	Provinces	Drainage	Geographic coordinates		N	Haplotypes
1	LUT	Lutshipuka. R.	Mwaba	Katanga	Luapula R.	10°14'S	28°20'E	3	CGA001
2	KAF	Kafubu. R.	Lubumbashi	Katanga	LuapulaR.	11°42'S	27°29'E	7	CGA001
3	KAS	Kasanka R.	Kasanka National park	Katanga	Luapula R.	-	-	3	CGA001, CGA002
4	LMW	Lake Mweru	-	Katanga	Lake Mweru	-	-	3	CGA001
5	LUP	Luapula R.	-	Katanga	Luapula R.	-	-	1	CGA003
6	KAD	Kando R.	Dianda	Katanga	Lualaba R.	10°45'S	25°50'E	5	CGA004
7	LUL	Lulua R.	Sandoa	Katanga	Kasaï R.	09°41'S	22°51'E	3	CGA005, CGA006, CGA007
8	DLB	Lualaba R.	Bukama	Katanga	Lualaba R.	09°11'S	25°51'E	10	CGA008, CGA009
9	DKA	Lake Kabwe	Butulu	Katanga	Lualaba R.	09°10'S	25°59'E	3	CGA008
10	DKL	Lualaba R.	Kalombo	Katanga	LualabaR.	08°18'S	26°19'E	5	CGA008, CGA010, CGA011, CGA12, CGA013
11	DMA	Lualaba R.	Malemba	Katanga	Lualaba R.	08°02'S	26°47'E	4	CGA008, CGA014
12	DKI	Lualaba R.	Kinkonja	Katanga	Lualaba R.	08°16'S	26°21'E	5	CGA008, CGA015
13	LKP	Lufira R.	Kapolowe	Katanga	Lufira R.	11°02'S	26°56'E	4	CGA016
14	LKI	Luambo R.	Kisunka	Katanga	Lufira R.	10°57'S	27°05'E	3	CGA016
15	LML	Mwera R.	Mulandi	Katanga	Lufira R.	10°50'S	26°57'E	5	CGA016
16	KYU	Lufira R.	Kyubo	Katanga	Lufira R.	09°31'S	27°02'E	5	CGA016
17	LUF	Lufira R.	Kapolowe station	Katanga	Lufira R.	11°21'S	26°46'E	10	CGA017, CGA018
18	LWA	Lwawe R.	Lipanda	Katanga	Lukuga R.	05°54'S	28°03'E	5	CGA019
19	KIS	Congo R.	Kisangani	Orientale	Congo R.	00°31'N	25°11'E	8	CGA008
20	KIN	Congo R.	Kinshasa	Kinshasa	Congo R.	04°18'S	15°20'E	10	CGA008

Table 2.1 (Continued)

Site #	Code	River or lake	Location	Provinces	Drainage	Geographic coordinates		N	Haplotypes
21	MBJ	Mbuji-Mayi R.	Mbuji-Mayi	Kasaï oriental	Kasaï R.	06°03'S	23°33'E	4	CGA008
22	LUB	Lubilanji R.	Mbuji-Mayi	Kasaï oriental	Kasaï R.	06°05'S	23°45'E	3	CGA008
23	RUG	Rugo R.	Rugo Bridge	Katanga	Lake Tanganyika	05°42'S	29°20'E	5	CGA020, CGA021
24	RUZ	Rusizi R.	-	-	Lake Tanganyika	-	-	3	CGA021
25	LKV	Lake Kivu	Bukavu	Kivu	Lake Kivu	02°29'S	28°51'E	2	CGA022, CGA023
26	UBA	Ubangi R.	-	-	Ubangi R.	-	-	1	CGA024

2.2.2. Alignment and phylogenetic analysis

A total of 118 cyt *b* sequences generated during this study were aligned together with published GenBank sequences of *C. gariepinus* from the Luapula (AF126823) and Ubangi system (AF235927) (Agnèse & Teugels, 2005) using the program ClustalX v. 2.0.11 (Thompson *et al.*, 1997; Larkin *et al.,* 2007). The alignment was further improved by eye. Nucleotide composition was determined using the software MEGA v. 4 (Tamura *et al*., 2007). Three haplotypes of *Clarias ngamensis* (AF23593, XXX and XXX) were selected as outgroup for the phylogenetic analyses. The appropriate evolutionary model was selected with jModelTest (Posada, 2008). According to the Akaike information criterion (AIC), the TPM1uf (Posada, 2009) + Γ model was the most suitable model for further analysis, with a gamma shape parameter of 0.11. Three complementary algorithms were used, namely neighbour-joining (NJ), maximum likelihood (ML) and Bayesian inference (BI) implemented respectively in MEGA v. 4., PAUP* v. 4.0 b10 (Swofford, 2002) and MrBayes v. 3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). For NJ and ML, the robustness of the nodes was assessed by bootstrap analysis with 1000 replicates. For neighbour-joining (NJ) the Tamura 3-Parameter (T3P; Tamura, 1992) was used with a gamma correction model because it is the only implemented model similar to TPM1uf + Γ available in MEGA v. 4. Sequence divergences were estimated using the same software and model. The general time reversible (GTR; Tavaré, 1986) model with a gamma distribution of site-specific rates was used in the Bayesian analysis because it is the next most complex model to TPM1uf + Γ available in MrBayes v 3.1.2. Posterior probabilities were calculated over 2.5×10^6 generations, while sampling the Markov Chain at a frequency of 100 generations. 25 % of the samples were discarded as "burn-in''.

2.2.3. Network analysis and nucleotide diversity

A median-joining network analysis was performed using the software NETWORK v. 4.5.0.1 (http://www.fluxus-engineering.com). Using the program DnaSP v. 5 (Librado

& Rozas, 2009), cyt *b* polymorphism was estimated as nucleotide (π: Nei, 1987) and haplotype diversity (h_d: Nei & Tajima, 1981) for each observed clade or cluster.

2.2.4. Divergence time between populations

There are no geological, hydrological or biological reference times for the evolutionary rate of catfish. With the parameters resulting from jModelTest, a likelihood-ratio-test (LRT) was carried out in TREE-PUZZLE v.5.2 (Schmidt *et al.*, 2002) in order to test the molecular clock hypothesis and assess whether the cyt *b* gene behaves like a strict clock. If so, divergence times can be estimated based on pairwise genetic distances from our sequence data. Agnèse & Teugels (2001b) used a mutation rate of 1.2 and 1.3 % per million years to estimate the divergence time between *C. gariepinus* of the Luapula system and *Bathyclarias* Jackson from Lake Malawi. In this study, it is proposed to use a range of molecular clock estimates for the cyt *b* fragment of 1 and 2 % per million years (Bermingham *et al*., 1997; Bowen *et al*., 2001; Domingues *et al*., 2005; Larmuseau *et al*., 2009) to calculate divergence times between the various clades and clusters.

2.3. Results

2.3.1. mtDNA haplotypes

The alignment of sequences was straightforward as there were no gaps and translation into amino acids did not indicate nonsense or stop codons. The sequence characteristics matched the general properties of the *C. gariepinus* cyt *b* gene (Agnèse & Teugels, 2001a), suggesting a functional mtDNA cyt *b* gene and not a nuclear pseudogene (Zhang & Hewitt, 1996). After alignment of the 120 sequences, 24 haplotypes were recognized. Our dataset of 551 bp contained a total of 52 variable sites, of which 35 were parsimony informative.

2.3.2. Phylogenetic analyses

Phylogenetic analysis of the 24 unique sequences revealed a Lulua-Kando clade and Congo clade, each well-supported by bootstrap and posterior probability values (Fig. 2.2). All other *C. gariepinus* haplotypes are probably to be considered clusters or clades but received low statistical support. The topologies of ML, NJ and BI trees were almost identical, with only small differences. The Congo clade united 9 haplotypes from the rivers Congo, Lualaba and Lwawe.

The Lulua-Kando clade included four haplotypes from the Lulua and Kando rivers. The Upper Lufira, Middle Lufira, Luapula-Mweru system, Tanganyika-Rusizi, Kivu and Ubangi clusters included, respectively, the two haplotypes from upper of river Lufira, one haplotype of Middle of river Lufira, three haplotypes of river Luapula, Kafubu and Luthipuka, two haplotypes of Lake Tanganyika and the Rusizi River, two haplotypes from Lake Kivu and one haplotype from the Ubangi system (Fig. 2.2 and Table A1.). A high sequence divergence (T3P distance) was observed between clades. The value of this distance ranged from 0.03 to 0.05 (Table A2.). Within the Congo clade, the substantial pairwise distance values ranged from 0.038 to 0.052.

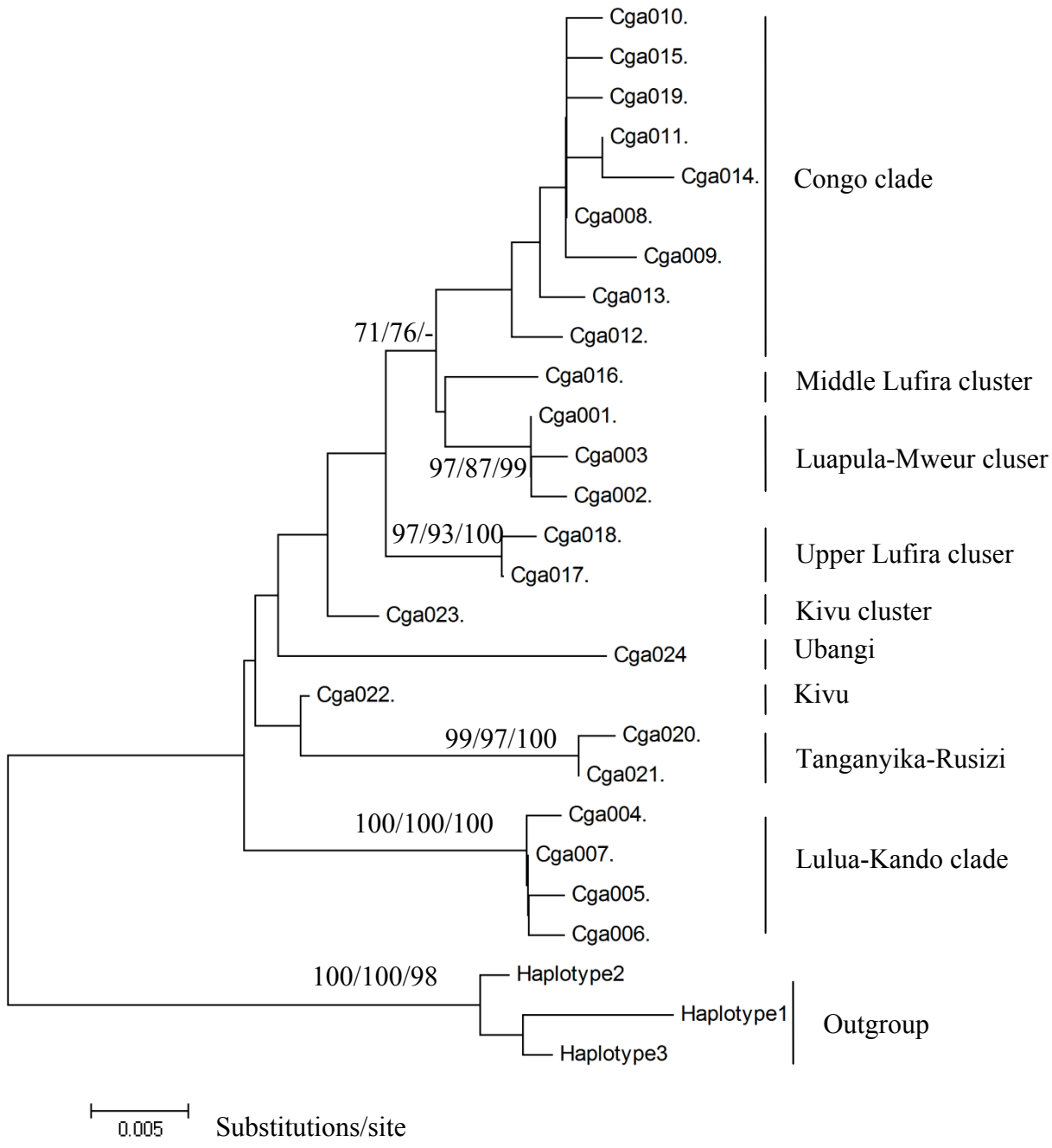


Figure 2.2. Neighbour-joining (NJ) tree of *Clarias gariepinus* with Tamura-3-P distance based on mitochondrial DNA *cytochrome b* sequences of the unique haplotypes. The numbers at the nodes represent the bootstrap values of neighbour-joining (NJ), maximum likelihood (ML) and posterior probabilities of Bayesian inference (BI) analyses. Values below 70% (Neighbour-joining and Maximum likelihood bootstrap support) or 80 % (Bayesian posterior probabilities) are marked with '-'. (See Table 2.1. for cyt b haplotype abbreviations).

2.3.3. Network and demographic analyses

Phylogenetic reconstruction using median-joining parsimony showed two different clades: the Congo clade and the Lulua-Kando clade (Fig. 2.2). The network of the Congo clade revealed four clusters (Fig. 2.3): the Congo-Lualaba cluster (green) containing the populations of the rivers Lualaba, Mbuji-Mayi, Lubilanshi, Lwawe and Congo; a Middle Lufira cluster (yellow) with the populations of the rivers Mwera and Lufira at Luambo and Kyubo; an Upper Lufira cluster (red); a Luapula-Mweru cluster (blue) containing the populations of the rivers Kafubu, Lutshipuka and Luapula, and Lake Mweru. Three clusters were not well resolved in the phylogenetic tree (Fig. 2.2) but are in the network analysis: the Kivu cluster (white) containing the populations of Lake Kivu, the Tanganyika-Rusizi cluster containing the populations of the rivers Rusizi and Rugo, and the Ubangi cluster containing the population of the Ubangi system (Fig. 2.3). The Congo-Lualaba cluster shows one main ancestral haplotype (Cga008), which is found at all locations except in the Lwawe River. The Congo-Lualaba group reveals a star-like pattern, probably reflecting a recent expansion (Fig 2.3). The Luapula-Mweru, Upper Lufira, Middle Lufira (Cga001, Cga017 and Cga016, respectively), which are found at all locations for each group. The genetic diversity was higher for the Congo clade (h_d = 0.752 and π = 0.008) than for the Lulua-Kando clade (h_d = 0.643 and π = 0.002). The Tanganyika-Rusizi cluster shows a higher genetic diversity (h_d = 0.571 and π = 0.001) than the other clusters (Table 2.2). The Ubangi and Kivu clusters include a low number of samples and it is hence impossible to determine the genetic diversity for these clusters.

2.3.4. Divergence time between populations

The LRT implemented in TREE-PUZZLE showed that a clock-like behavior of the sequences should not be rejected at a significance level of 0.05. Therefore the divergence times were estimated at two mutation rates. A divergence time of 1.1 and 2.2 MYA respectively was obtained for the 2 and 1% of evolution rate between the Congo and Lulua-Kando clade (Table 2.3). The divergence time between populations

of the Congo Central Basin and the Kamalondo Depression was respectively 0.1 and 0.2 MYA.

Table 2.2. Genetic diversity of the populations of *Clarias gariepinus* based on mitochondrial cytochrome *b* data: number of specimens (N), number of haplotypes (N_H), haplotype diversity (h_d) and nucleotide diversity (π).

Unit	**N**	**N_H**	**h_d (S.D.)**	**π (S.D.)**
Total data set	120	24	0.822 (0.027)	0.0140 (0.002)
Lulua-Kando clade	8	4	0.643 (0.184)	0.0020 (0.001)
Congo clade (Congo-Lualaba cluster)	59	9	0.413 (0.081)	0.0030 (0.015)
Middle-Lufira cluster	17	1	0.118 (0.101)	0.0010 (0.001)
Upper-Lufira cluster	10	2	0.200 (0.154)	0.0004 (0.001)
Luapula-Mweru cluster	15	3	0.257 (0.142)	0.0014 (0.002)
Tanganyika-Rusizi cluster	8	2	0.571 (0.094)	0.0010 (0.001)

Table 2.3. Divergence time in million years (MYA) between different groups of *Clarias gariepinus* of the Congo Basin for the mutation rate of 1 and 2 %.

Comparison between groups²	**Divergence time (MYA)**	
	2%	**1%**
Congo /Lulua-Kando clades (4.4% T3P)	1.100	2.200
Congo- Lualaba/Upper-Lufira clusters (2.1% T3P)	0.525	1.050
Congo- Lualaba/Middle-Lufira clusters (1.5% T3P)	0.375	0.750
Congo- Lualaba/Luapula-Mweru clusters (1.6% T3P)	0.400	0.800
Congo- Lualaba /Tanganyika-Rusizi clusters (5% T3P)	1.250	2.500
Congo- Lualaba/Ubangi clusters (4.5% T3P)	1.125	2.250
Congo- Lualaba/Kivu clusters (2.2% T3P)	0.550	1.100
Congo Central Basin/Kamalondo depression populations (0.4% T3P)	0.100	0.200

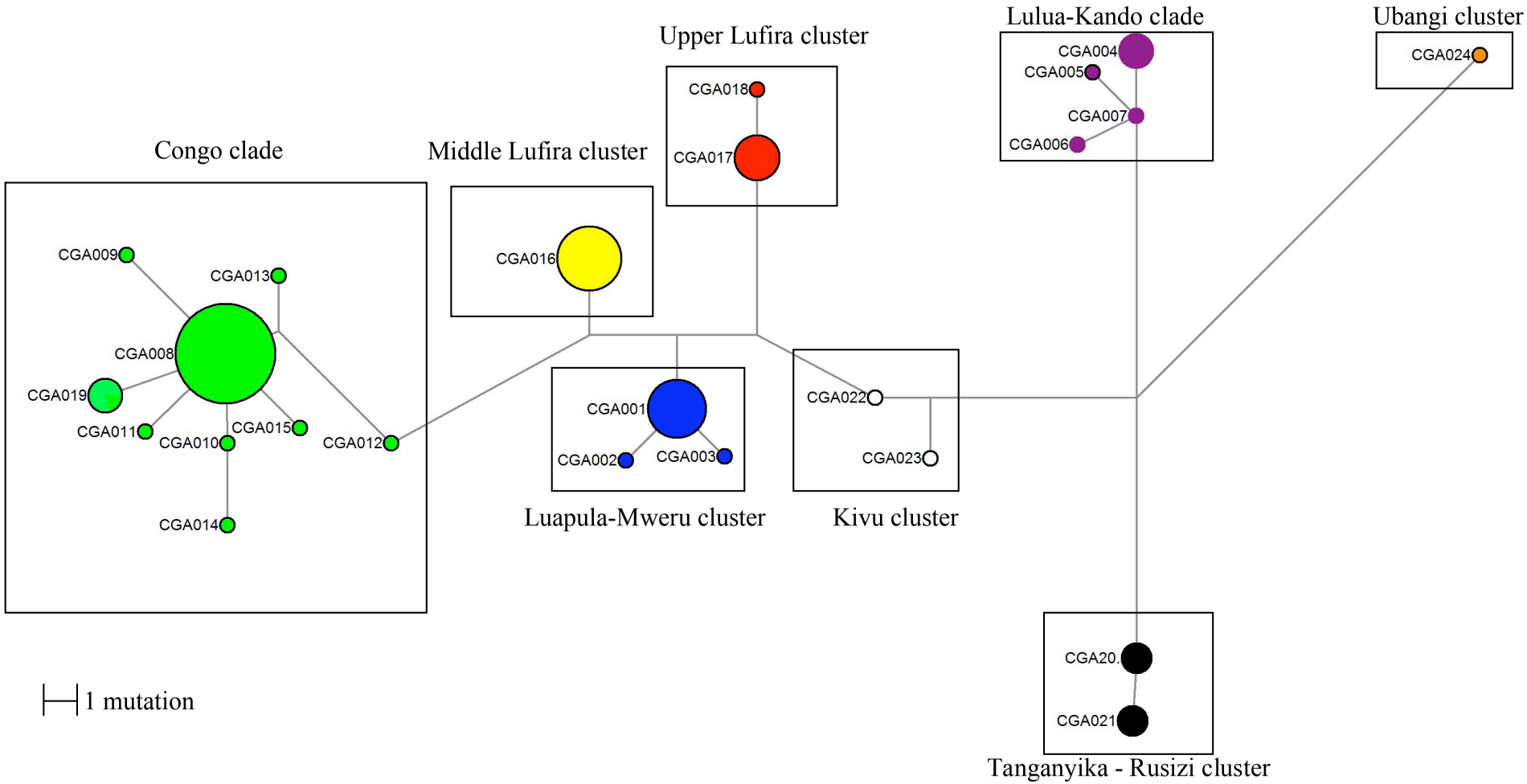


Figure 2.3. Median-joining network of mitochondrial DNA cytochrome *b* haplotype in Congolese *Clarias gariepinus* populations. The size of the circles is proportional to number of Catfish sharing that haplotype. See Table 2.1 for the code of haplotype.

2.4. Discussion

2.4.1. Diversity and lineages

From this study appears a widespread group of *C. gariepinus* inhabiting the Congo Basin and six additional clusters and one clade inhabiting the uplands at the periphery of the Basin. A most significant finding is the moderate diversity (h_d=0.413 and π=0.0030) of the populations at the cyt *b* locus in the Congo Basin proper (Cuvette Centrale). Most other genetic studies on catfish, using a range of genetic markers, suggested a high local diversity (π=0.0660) (Agnèse & Teugels, 2001b, 2005; Giddelo *et al*., 2002; Arndt *et al*., 2003; Mwita & Nkwengulila, 2008). This is remarkable as the Congo Basin is thought to belong to its ecological core, given its common occurrence and high abundance (Teugels, 1986). The vast region corresponding to the Congo Basin is occupied by a single clade (Congo). Southwest of the Congo Basin lives a second clade, namely the Lulua-Kando clade of western Katanga. Few additional information on local catfish diversity could be found but Teugels (1986) mentions its common occurrence in the region. Katanga province is considered to harbour an important although poorly described biodiversity (J. Snoeks, pers. comm.). Broadley & Cotterill (2004) and Cotterill (2006) mention its herpetological and ornithological significance.

Three clusters occupy the south (Upper Lufira; Middle Lufira; West Zambian Congo branches including the Luapula River and Lake Mweru). The centrally located cluster of the Congo Basin – Kamalondo Depression (Upemba region) occupies by far the vastest region from the southern Kamalondo Depression (Lualaba River - 575 m above sea level) downstream up to Kinshasa (270 m above sea level), a distance of at least 2500 km. One very common ancestral haplotype is found at all locations (Kinshasa, Kisangani and Kamalondo depression) except in the Lwawe River (Lukuga basin). A set of rapids on the Congo River (Stanley Falls and Portes de l'Enfer) do not seem to have influenced the pattern to a large extent. It was therefore suggested that the ancestor of *C. gariepinus* colonizing the Congo Basin originated from the Kamalondo Depression. The fauna of the Kamalondo Depression resembles the Congo Basin and shows close historical ties (Poll, 1976).

The Upper Lufira cluster (Red) occupies the Katanga plateau (1100 m above sea level) and is separated from the Middle Lufira (800 m above sea level) by the barriers such as Mwadingusha and Koni water falls. Fish faunas below and above the water falls are distinct (Magis, 1961; Poll, 1963). As the Upper Lufira and Upper Kafue (Zambezi basin) share a common history of river captures (Stankiewicz & de Wit, 2006), most likely the Upper Lufira cluster has communalities with the Zambezi clade.

The West Zambian Congo branches of the Luapula – Lake Mweru system (blue) (900 m above sea level) is separated from the Kamalondo Depression (560 m above sea level) by the Johnson water falls on the Luapula River and the Mututa water falls on the Luvua River (Banister, 1986). The region is well documented for its influence from the Upper Zambezi River (Poll, 1976; Thieme *et al.*, 2005). Historically the Luapula-Mweru system has been part of the Zambezi system during the early Tertiary (Skelton, 1994) and showed a link between the Upper Zambezi and Congo river untill fairly recently. Katongo *et al*. (2005) reported a strong impact of the local geography on the differentiation of haplochromine cichlids in the Chambeshi-Bangwelu and Mweru populations. Also Katongo *et al.* (2005) and Joyce *et al.* (2005) showed that the expansion of haplocromines cichlid towards southern Africa originated in the central Congo River, by colonising first the Upper Congo-Luapula system, and then entering the Zambezi system further south via the Zambezi - Congo watershed. This scenario is supported by the positioning of *Serranochromis* Regan (from Lake Mweru in the Upper Congo) as a sister group to the genera *Sargochromis* Regan and *Pharyngochromis* Greenwood of the Zambezi and several other haplochromine cichlids (Salzburger *et al*., 2005).

The catfish cluster of Lake Tanganyika–Rusizi River is represented by two haplotypes, not that well resolved phylogenetically, but clearly distinct in the network. It points to the distinct history of the lake region, well known for its radiation of haplochromine cichlids (Koblmüller *et al*., 2008). The connection between the Congo Basin and Lake Tanganyika via the overflow to the Lukuga valley has been strongly influenced by tectonics (rifting) and lake level, itself related to Pleistocene climate cycling (Cohen *et al*., 1993). Lake Tanganyika had a last low stand at – 600 m some 18,000 Y BP (Johnson at *al*., 1996). Giddelo *et al*. (2002) attributed the same Rusizi catfish to the

Eastern African clade. Lake Kivu harbours distinct haplotypes, evolutionary more affiliated with the Upper Lufira and Luapula - Mweru cluster. There is currently no hydrographic link between these systems, but in the past Lake Kivu flowed to Victoria Lake (Verheyen *et al*., 2003); the flow reversed to the Congo Basin under influence of rifting in the Holocene (Lévêque, 1997).

The Ubangi River sample, consisting of just one individual analyzed by Agnèse & Teugels (2001b), represents its own cluster, resolved in the network but not in the phylogeny. It is putatively linked to the Northern African clade. The fauna of northern Congo has affiliations with the Nilo-Sudanic fauna (Bailey, 1986), exemplified by amongst others the cichlid fishes *Sarotherodon galilaeus* Linneaus 1758, subspecies *galilaeus* and *Tilapia zillii* Gervais 1848 (Thys van den Audenaerde, 1964). It is clear that more extensive sampling in the region should clarify the faunal exchange which might have happened through river capture.

The two clades and six clusters of *C. gariepinus* are clearly geographically separated in the Congo Basin. Such evidence reflects the dynamics of specific river systems as dispersal routes for fishes, which is shaped by temporary connections, river captures and boundary displacements (Bermingham & Martin, 1998; Hubert *et al*., 2007). The 18 major freshwater ecoregions of the Congo Basin based on the distribution of freshwater fishes (Thieme *et al*., 2005) provide an interesting source of information to understand some of the environmental drivers of differentiation and adaptation (Moodley & Bruford, 2007). For example, Thieme *et al*, (2005) proposed that the Upper Lualaba together with the Kamalondo Depression represent a single ecoregion. This study assigns a molecular assemblage specific to each region, suggesting distinct evolutionary histories of at least catfish, and maybe other taxa. The Congo Basin proper has been split in several ecoregions, while catfish shares haplotypes between these regions. It suggests strong historical connectivity of catfish between ecoregions.

2.4.2. Pleistocene dynamics and the role of refuges

The historical population dynamics of *Clarias gariepinus* in the Congo Basin show evidence for Plio-Pleistocene divergence; our data point to two distinct clades and a

minimum of four clusters within one of these clades. A more advanced interpretation which incorporates the literature (Rognon *et al,* 1998; Agnèse & Teugels, 2001b; Giddelo *et al.* 2002) would suggest four clades (northern Africa – Ubangi; eastern Africa – Lake Kivu, Lake Tanganyika and Rusizi; southern Africa – West Zambian Congo; central Africa – all other taxa). It would require the reassignment of some of the groups to the clade status, a proposal which awaits confirmation. In addition, a relatively recent range expansion of the Congo-Lualaba and Middle Lufira groups contrasts with more stable populations in many of the clusters neighboring the Congo Basin. It is for example remarkable how a single group occupies the Central Congo Basin. One interpretation is that the range expansion accompanied an historical extension of suitable habitat (see later).

Major drivers of divergences and expansions in aquatic systems have been traced back across the globe to the historical dynamics of climate, geology and hydrology (Hewitt, 2004). For terrestrial systems the evolution of the vegetation in response to global climate change represents an additional factor influencing the evolution of species (Maley, 2001). It has been traced in the historical patterns of African mammals and hominids (deMenocal, 2004; Maslin & Christensen, 2007; Trauth et al. 2009). The Congo Basin, centrally located and one of the most significant ecoregions in Africa, represents no exception. Vast changes of its river systems, geological uplift in the East and South, rifting in the East and considerable climate change throughout the late Tertiary and Quaternary have left their traces, some of which have been picked up in this study.

Remarkable is that historical divergence occurred in waves, which has been linked to key divergences across faunal taxa in concordance with Pleistocene climate dynamics (deMenocal, 2004). Some 2.8 MYA the grasslands expanded to the detriment of the forested zones; 1.7 MYA another expansion affects the coverage of grasslands; 1.0 MYA begins a series of 100 kYA glacial cycles as traced back in many taxa (e.g. bushbuck - Moodley & Bruford, 2007; elephant - Comstock *et al*, 2002; Johnson *et al*., 2007; forest monkeys - Tosi, 2008; shrews - Quérouil *et al.*, 2003; other mammals - De Vivo & Carmignotto, 2004; Hewitt, 2004). Hydrological changes of extremely large amplitude, including dry periods in the basins surrounding the Congo Basin

during the Last Glacial Maximum (LGM; Gasse, 2000) might (wrongly) suggest that the catfish of the Congo Basin are relicts. Arndt *et al*. (2003) showed that the population fragmentation of catfish in northern Africa is consistent with wet/dry cycling in the Sahara during the Pleistocene. The consistent availability of river and wetland habitat in the Congo Central Basin during the LGM has guaranteed that populations of *C. gariepinus* remained viable. A fluvial refuge seems to have persisted during the arid periods of the late Pleistocene (Beadle, 2001; Clifford *et al*., 2004; Hughes & Hughes, 1992).

Changes in river hydrology are closely associated with the geological dynamics of the region. Two main processes have been playing a role, one being the African Superswell of eastern and southern Africa responsible for the high altitude and leading to uplifting (Pasyanos & Nyblade, 2007). With this process is associated rifting between the African plate and Somali plate including the East African rift valleys. The uplifting and rifting have led to fragmentation and river captures in the eastern and southern range of the Congo Basin. As the process is ongoing since the Tertiary, populations of *C. gariepinus* have become fragmented in the West Zambian Congo, Lake Tanganyika and Kivu, the Upper and Middle Lufiira clusters, and Lulua-Kando clade. A second, more recent, process is the drainage of the Central Congo Basin (Beadle, 1981). The lowering of the water level led to the current situation of a dense network of rivers on the backbone of the immense Congo River. *C. gariepinus* is a demersal shallow water fish and hence did not occupy the pelagic zone of the lake. But once drained, it invaded from the Upper Lualaba (Kamalondo Depression) downstream an immense habitat, covering several ecoregions (Thieme *et al.,* 2005). The general appearance of a single cluster from the Upper Lualaba up to Pool Malebo in the Lower Congo is the testimony of this process.

We propose the following interpretation of the historical dynamics of *C. gariepinus* in three phases. In the late Pliocene clades diverged over a vast region of Central Africa linked to the African Superswell and rifting. Cluster divergence in the Congo Basin occurred during the Pleistocene, when the relief to the east and south of the Central Congo Basin was conducive to isolation and micro-evolution. Upon emptying of the

Central Congo Basin, *C. gariepinus* expanded from the Kamalondo region in a northern and than western direction across the huge Congo Basin.

The question arises now what the evidence is, especially for the last phase, in other aquatic and terrestrial taxa? The few data available on the Congo Basin point first of all to the high affinity between the fish fauna of the Central Congo Basin and the upper Lulalaba (Kamalondo Depression) and Lufira (Poll, 1963). Poll proposes a scenario of a recent connection between the Kamalondo Depression and the Central Basin. Eaton *et al*. (2009) observed that the large Basin (rain forest) between the Ituri forest (Democratic Republic of the Congo) and Lake Télé Community Reserve (Republic of the Congo) is occupied by a single clade of the dwarf crocodile (*Osteolaemus* sp.). Evidence from terrestrial taxa such as simians (gorilla, chimpanzee and bonobo) and elephant (Anthony *et al*., 2007; Johnson *et al.,* 2007) point to population dynamics in relation to forest coverage and fragmentation due to river barriers.

2.4.3. Implications for aquaculture and conservation

Aquaculture

A large historical diversity and divergence harbours a great adaptive potential and hence represents an important source for the selection of aquaculture strains. Huisman & Richter (1987) collected the Ubangi strain, which has experienced a process of domestication, although with some crossing with the Israeli strain. It is used for aquaculture in The Netherlands and has been reported as the Dutch strain in Africa (Anene & Tianxiang, 2007) and SE Asia (Wachirachaikarn *et al.*, 2009). Another attempt has been made to domesticate *C. gariepinus* in South Africa (Van der Bank *et al*., 1992). Some strains have been used occasionally for local aquaculture in the Congo Basin (e.g. Lubumbashi – pers. observ.; Kisangani – H. Gevaert, pers. comm; Rwanda – E. Rurangwa, pers. comm.), but have not been domesticated as such. Catfish is regionally popular as judged from the good price it fetches on local markets. Although the commercial production is growing in Europe, Asia and Latin America, Africa's contribution remains small and stagnates (Hogendoorn *et al*., 1983; Ducarme & Micha, 2003).

Here is an opportunity for detailed strain characterisation and phenotyping of growth, ecophysiology, disease resistance and stress. The rational utilization of genetic resources requires the assessment of their genetic and phenotypic diversity. Given the large historical divergence between strains, cultured stocks should be maintained without transfer between ichtyofaunal provinces. It will prevent genetic contamination of the native gene pool (Doupé & Lymbery, 2000).

Conservation

Wildlife in the Congo Basin is extremely rich but at the same time very fragile. It is threatened by many anthropogenic factors: population growth, infrastructure, extraction of minerals, logging, agriculture, fishing, pollution from cities and mining, and climate change (Thieme *et al*, 2005). *C. gariepinus* is clearly not a rare species, but a good indicator of historical diversity and a rich source of protein. Especially the Katanga region appears as a hotspot with its known endemisms of mammals such as the antilope *Kobus anselli* (Cotterill, 2005) and fishes such as *Parakneria lufirae* Poll and the cichlid *Oreochromis salinicola* Poll (Thys van den Audenaerde, 1964; Poll, 1976). Wildlife management, if implemented, can rely on catfish as a useful indicator. It is a robust fish adapted to several categories of waters with a high fecondity and adaptation to several types of waters (Teugels, 1986). His disappearance may point to overfishing and pollution.

2.4.4. Perspectives

The phylogeography of catfish in the Congo Basin shows the presence of four clades. But what is the link with the populations of neighboring basins, such as the Nile (northern Africa clade) and the Zambezi (southern Africa clade)? A second interesting question would address the connectivity on a contemporaneous scale within the Congo Basin. What is the realized link between catfish populations over a distance of more than 2500 km? Finally, the phenotypic characterization of the aquaculture potential of the various clades and clusters might prove helpful for the aquaculture of catfish at various altitudes and under various environmental conditions.

Acknowledgements

We thank A. Kankonda (Université de Kisangani) and J.-P. Marquet (PRODEPAAK/ BTCCTB) for help with field sampling. D. Schaerlaekens helped with the scoring of the genotypes, J. Raeymaekers and L. Mbaya helped with the GIS. Funding was provided to ACM by the Belgian Technical Cooperation, the K.U.Leuven to MHDL and the Research Foundation-Flanders (FWO-Vlaanderen) to MPMV. B. Hellemans helped with molecular biology.

Appendices

Table A1. Haplotype distribution of cytochrome b haplotypes of Clarias gariepinus; the columns refer to a single population, while the rows refer to distribution of the haplotypes. The haplotypes are indicated by the codes in Table 2.1.

	LUT	KAF	KAS	LMW	LUP	KAD	LUL	DLB	DKA	DKL	DMA	DKI	KAP	LKI	MUL	KYU	LUF	LWA	KIN	KIS	MBJ	LUB	RUG	RUS	LKV	UBA
CGA001	3	7	2	3																						
CGA002			1																							
CGA003					1																					
CGA004						5																				
CGA005							1																			
CGA006							1																			
CGA007							1																			
CGA008								9	3	1	3	4							10	8	4	3				
CGA009								1																		
CGA010										1																
CGA011										1																
CGA012										1																
CGA013										1																
CGA014											1															
CGA015												1														
CGA016													4	3	5	5										
CGA017																	9									
CGA018																	1									
CGA019																		5								
CGA020																							1			
CGA021																							4	3		
CGA022																									1	
CGA023																									1	
CGA024																										1

Table A2. Pairwise distances according to the Tamura 3 P model between the cytochrome *b* haplotypes of the populations of *Clarias gariepinus* in the Congo basin. For haplotype abbreviations see Table 2.1.

	GA001	GA002	GA003	GA004	GA005	GA006	GA007	GA008	GA009	GA010	GA011	GA012	GA013	GA014	GA015	GA016	GA017	GA018	GA019	GA020	GA021	GA022	GA023
GA002	0.002																						
GA003	0.002	0.004																					
GA004	0.039	0.043	0.042																				
GA005	0.039	0.043	0.042	0.004																			
GA006	0.039	0.043	0.042	0.004	0.004																		
GA007	0.036	0.039	0.039	0.002	0.002	0.002																	
GA008	0.012	0.014	0.014	0.044	0.044	0.044	0.041																
GA009	0.016	0.019	0.019	0.052	0.052	0.052	0.048	0.004															
GA010	0.014	0.016	0.016	0.048	0.048	0.048	0.044	0.002	0.006														
GA011	0.014	0.016	0.016	0.048	0.048	0.048	0.044	0.002	0.006	0.004													
GA012	0.014	0.016	0.016	0.041	0.041	0.041	0.038	0.006	0.010	0.009	0.008												
GA013	0.016	0.019	0.019	0.052	0.052	0.052	0.048	0.004	0.008	0.006	0.006	0.006											
GA014	0.019	0.021	0.021	0.051	0.051	0.051	0.047	0.006	0.010	0.008	0.004	0.012	0.010										
GA015	0.014	0.016	0.016	0.048	0.048	0.048	0.048	0.002	0.006	0.004	0.004	0.008	0.006	0.008									
GA016	0.010	0.012	0.012	0.036	0.036	0.036	0.036	0.014	0.019	0.016	0.016	0.008	0.010	0.021	0.016								
GA017	0.015	0.017	0.017	0.036	0.036	0.036	0.036	0.016	0.021	0.019	0.019	0.019	0.021	0.024	0.019	0.017							
GA018	0.017	0.019	0.020	0.039	0.039	0.039	0.039	0.019	0.019	0.021	0.021	0.021	0.024	0.026	0.021	0.019	0.002						
GA019	0.014	0.016	0.016	0.048	0.048	0.048	0.048	0.002	0.006	0.004	0.004	0.008	0.006	0.008	0.004	0.016	0.019	0.021					
GA020	0.043	0.046	0.046	0.043	0.043	0.043	0.042	0.048	0.056	0.052	0.052	0.052	0.056	0.055	0.052	0.047	0.040	0.044	0.052				
GA021	0.040	0.040	0.043	0.039	0.039	0.040	0.039	0.045	0.052	0.049	0.049	0.049	0.052	0.052	0.049	0.043	0.037	0.040	0.049	0.002			
GA022	0.014	0.019	0.019	0.027	0.027	0.027	0.027	0.021	0.027	0.024	0.024	0.024	0.027	0.026	0.024	0.019	0.019	0.022	0.024	0.019	0.017		
GA023	0.012	0.015	0.014	0.027	0.027	0.027	0.027	0.016	0.021	0.019	0.019	0.019	0.021	0.024	0.019	0.015	0.015	0.017	0.019	0.025	0.022	0.004	
GA024	0.041	0.045	0.044	0.047	0.047	0.047	0.047	0.037	0.043	0.040	0.040	0.040	0.043	0.043	0.040	0.045	0.032	0.035	0.040	0.045	0.041	0.029	0.029

Table A3. Variable sites of the 551 bp fragment of cytochrome *b* among the 24 haplotypes recognized in *Clarias gariepinus*.

Polymorphic site	.	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	5	5	5	
Haplotype		4	4	7	8	9	0	1	2	3	3	4	5	5	6	6	8	9	1	2	3	5	5	6	7	8	8	9	0	0	1	1	2	2	3	4	6	7	9	1	3	4	4	5	5	6	6	7	8	1	1	4	**Accession**
		0	3	9	5	1	6	8	4	0	6	5	1	4	3	9	0	4	4	4	8	3	9	8	1	0	6	2	1	4	3	6	5	8	1	0	4	7	7	5	9	2	5	2	6	0	3	8	1	4	7	7	**Number**
CGA001		C	T	T	A	A	A	C	A	C	C	C	T	C	C	C	T	G	C	C	A	T	C	T	C	C	C	T	C	A	A	C	C	C	C	C	C	C	A	A	C	C	G	C	A	T	A	C	C	C	C	
CGA002																																										T										
CGA003																		T																																T	**AF126823**
CGA004												C	T					T	A			C		C				G					T	T			T		G				G	C		T	T		T	
CGA005										T		C	T					T	A			C		C				G					T	T			T		G					C		T	T		T	
CGA006												C	T			C		T	A			C		C				G					T	T			T		G					C		T	T		T	
CGA007												C	T					T	A			C		C				G					T	T					G					C		T	T			
CGA008								A					T								C				G				G															C						
CGA009		T						A					T					T			C				G				G													A			C						
CGA010								A					T								C		T		G				G															C						
CGA011								A					T								C				G				G											G				C						
CGA012								A					T								C				G			G									T						C						
CGA013								A				T	T								C				G				G									T						C						
CGA014								A					T								C			G	G				G											G				C			T			
CGA015								A					T								C				G			T	G															C						
CGA016												T	T												G													T						C						
CGA017													T			T				G					T									T									C						
CGA018													T			T				G					T									T								A		C						
CGA019								A					T								C			T	G				G															C				T		
CGA020			C	C		G	G				T		T	T										C							T	T	T									C		T	T				
CGA021			C	C		G	G				T		T	T										C							T	T	T									C	G	T	T				
CGA022				C									T	T										C									T									C		T					
CGA023				C									T											C									T	T									C			T				
CGA024				C	T				T				T								C			C	T	C			A	T		T						T	T				C	G	T	T			**AF235927**

Chapter 3

Genetic patterns of fishes in large rivers: old units and recent dispersal of *Clarias gariepinus* in the Congo Basin

Chocha Manda A., Larmuseau M., Vanhove M., Micha J.-C., Volckaert F.A.M.

Manuscript in preparation

Abstract

The evolution of fishes in large rivers is influenced by geological and climatic past events. In this chapter the dispersal capacity and genetic structure of the catfish *Clarias gariepinus* in the Congo Basin was analyzed with microsatellite markers to understand its contemporary evolution**.** In total 280 catfish from eight sites scattered across the basin were genotyped at seven microsatellites. Genetic diversity values were comparable except at Kapolowe Station (low value). The same Kapalowe Station and Dianda deviated significantly from Hardy-Weinberg equilibrium. Genetic differentiation between samples was high, except for those collected on the Congo River between Bukama and Kinshasa, a distance of at least 2500 km. Catfish inhabiting the various drainages of the upper Congo Basin in the Katanga province seem to be isolated, while fish living in the Congo River show a high level of connectivity over very long distances. These results agree with previous findings which focused on a historical perspective based on mitochondrial DNA. Of the three hypotheses tested to explain the origin of this accumulation of diversity, two were retained, namely the large dispersal hypothesis (Congo River) and the refuge hypothesis (Upper Congo; Katanga). The Katanga region is confirmed as an important region of biological diversity.

3.1. Introduction

Water is widely regarded as an essential natural resource and rivers represent a major component of its cycle. Some of the largest rivers worldwide are found in the intertropical region; Amazon, Congo and Mekong cover 38 % of all tropical basins (Latrubene *et al.,* 2005). Their relatively long and stable biological history makes that they are characterized by a highly diverse fish fauna to the degree that some sites represent hotspots of biodiversity (Myers *et al*., 2000; Küper *et al*., 2004) and belong to some of the last true wilderness areas (Cincotta *et al*., 2000). The Amazon Basin with more than 2500 neotropical species of fishes is the largest freshwater basin in the world (Hubert & Renno, 2006) followed by the Congo Basin (more than 1000 species) (Thieme *et al.,* 2005) and the Mekong Basin (more 1000 species) (Dudgeon, 2000). The question arises how this diversity has accumulated with time. This might have happened gradually over a long period (the Museum model), or relatively recently either through isolation (refuge function) or long-distance dispersal (Lévêque, 1997; Hubert *et al*., 2007). At the same time, the aquatic fauna of the large rivers has been influenced continuously by numerous minor and major events (climate fluctuations: Trauth *et al*., 2009; river captures: Lévêque, 1997; dispersal across barriers: Avise, 2000).

The typical biotas of the Congo basin have developed in an endorreal context throughout the Pliocene (5.3- 2.5 MYA) and Pleistocene (2.6 MYA - 12 kYA). There is a good understanding of peripheral biota such as Lake Tanganyika, which emerged some 9-12 million years ago (Lévêque, 1997), unlike the dynamics of the fauna of the Congo Basin (Thieme *et al.,* 2005; Lévêque, 1997). This raises not only the question to what degree species and their communities have evolved, but also to what degree intraspecific patterns have emerged. The analysis of the genetic structure of populations provides a suitable approach to study their evolutionary history. For example, ungulates have been the focus of numerous studies on the African continent (Brown *et al*., 2007; Lorenzen *et al*., 2006; Nersting and Arctander, 2001; Hewitt, 2004), however very few include material from the Congo basin (Hewitt, 2004). This has hampered a generally accepted scenario on the regional faunal dynamics. However, Chocha *et al.* (Chapter 2) noticed that the catfish *Clarias gariepinus*

occupies distinct evolutionary significant units at the periphery of the Congo basin, in contrast to the central basin which seems to have been colonized relatively recently. This was attributed to access to a vast new habitat following the drainage of the Central Basin after its connection with the Atlantic Ocean.

In this follow-up study to Chocha *et al.* (Chapter 2), the dispersal capacity and genetic structure of the catfish *Clarias gariepinus* was analyzed to understand the evolution of the contemporary ichthyofauna in the Congo Basin**.** Microsatellite markers provide an alternative perspective on the historic patterns and hence are suitable to complement the mitochondrial DNA patterns studied in Chapter 2.

3.2. Materials and methods

3.2.1. Population sampling

We collected a total of 280 adult *Clarias gariepinus* (Burchell, 1822) (Clariidae, Teleostei) individuals from eight sites in the Congo Basin, spread over the provinces of Katanga, Orientale and Kinshasa in the Democratic Republic of Congo between September 2006 and January 2010 (Fig. 3.1, Table 3.1). We purchased live fish from local fishermen, identified them on site and collected barbel clips for preservation in pure ethanol (100%).

3.2.2. Microsatellite genotyping

Genomic DNA was extracted from barbel clips using the NucleoSpin®96 Tissue Kit (Macherey-Nagel GmBH, Düren, Germany) following the manufacturer's instructions. A high-throughput protocol for the genotyping of seven microsatellite loci was developed. Five loci (Cga01, Cga02, Cga05, Cga09 and Cga11) were selected by testing and redesigning microsatellite markers previously isolated from *Clarias gariepinus* (Galbusera *et al*., 1996) and two loci (Cba 19 and Cba 20) isolated from *C. batrachus* (Yue *et al.,* 2003). All loci were amplified with the Qiagen Multiplex PCR kit (Qiagen, Venlo, The Netherlands). The following conditions of protocol of Qiagen Multiplex PCR kit were used for amplification (Table 3.2).

Multiplex PCR products were run on an ABI 3130-Avant capillary sequencer (Applied Biosystems) together with the internal size standard 500LIZ (Applied Biosystems). Fragment analysis was conducted using the software Genemapper v. 4.0 (Applied Biosystems). Potential genotyping errors, such as allelic dropouts, stuttering or null alleles, were analysed using MICRO-CHECKER v. 2.2.1 (Van Oosterhout *et al*. 2004). We detected scoring errors for the loci Cga 09 and Cba 19 and excluded them from the final analyses.

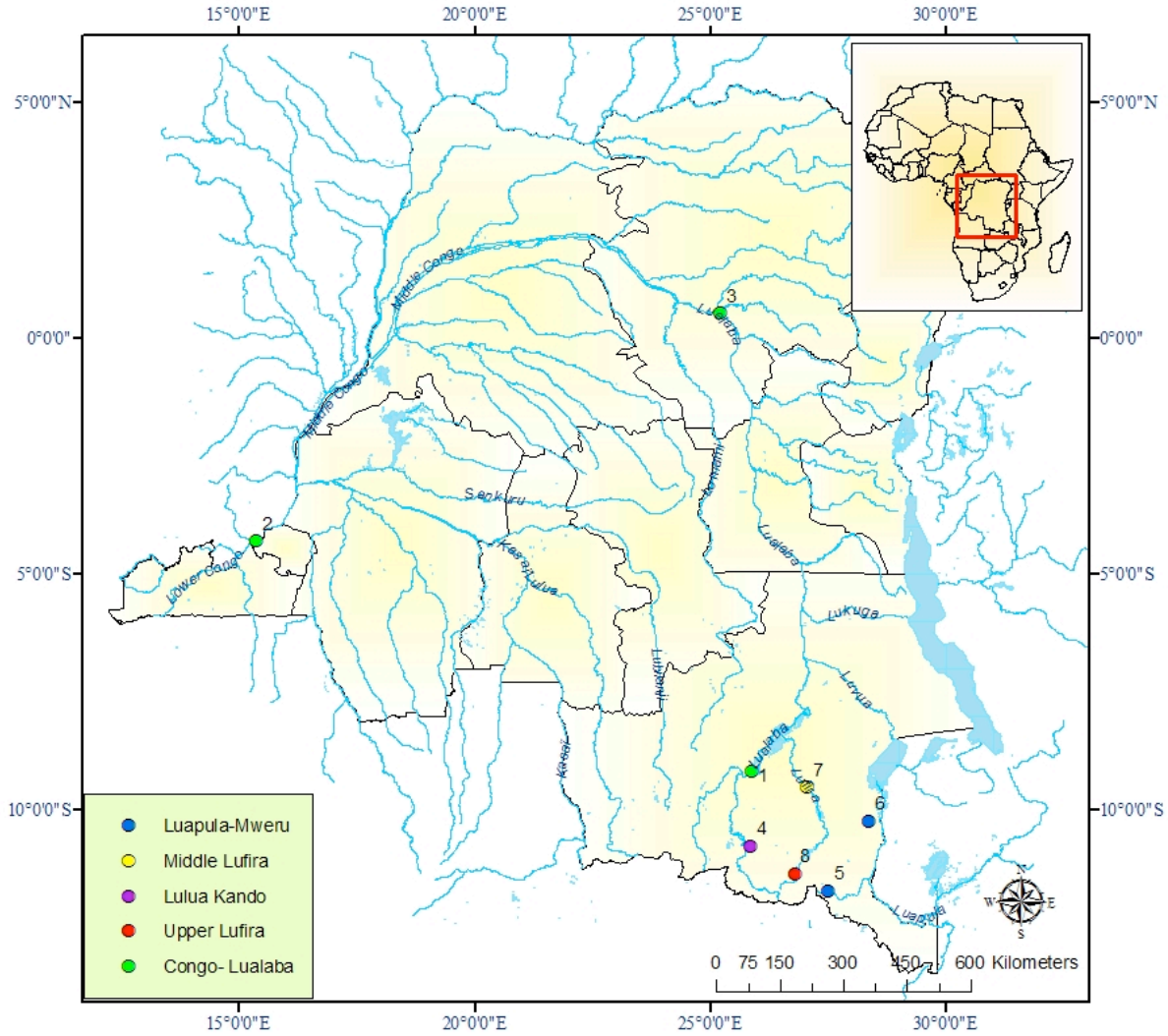


Figure 3.1. Geographical location of the samples of *Clarias gariepinus*. See Table 3.1 for sample codes.

Table 3.1. Listing of the area code, river, sampling site, geographical coordinates and number of samples per site (n).

Site	Code	Major drainage	River	Location	Provinces of D.R. of CONGO	Geographic coordinates		Altitude (m)	n
1	DLB	Lualaba	Lualaba R.	Bukama	Katanga	09°11'S	25°51'E	572	47
2	KIN	Congo	Congo R.	Kinshasa	Kinshsa	04°18'S	15°20'E	210	20
3	KIS	Congo	Congo R.	Kisangani	Orientale	00°31'N	25°11'E	400	38
4	KAD	Lualaba	Kando R.	Dianda	Katanga	10°45'S	25°50'E	1254	26
5	KAF	Luapula	Kafubu R	Lubumbashi	Katanga	11°42'S	27°29'E	1192	25
6	LUT	Luapula	Lutshipuka. R.	Mwaba	Katanga	10°14'S	28°20'E	976	29
7	KYU	Lufira	Middle Lufira R.	Kyubo	Katanga	09°31'S	27°02'E	800	48
8	LUF	Lufira	Upper Lufira R.	Kapolowe Station	Katanga	11°21'S	26°46'E	1129	47

Table 3.2. Characterization of seven *Clarias gariepinus* and *C. batrachus* microsatellite primer sets, including locus name, primer sequences, Genbank Accenssion number, specific anneling temperature, size-range of PCR. TD: Mutiplex Tuch down

PCR plex	Locus (GenBank no.)	Origin	Primer sequences (5' -> 3')	Annealing temp (°C)	Number of cycles	Primer Concentration (µM)	Size range (bp)	Number of alleles
Multiplex 1	Cba19 (AY169271)	*Clarias batrachus*	F: CAGGGCTAAATTACCCATAATCA R: GGCATGTGTTATAACATGTGAGG	58-52 (TD) 52	7 28	0.05	210-240	3
	Cba20 (AY169272)		F: GAAACACGCCATCATGCCTAATA R: CCAAACGGAGCGGACAGG			0.05	100-140	2
Multiplex 2	Cga01 (U30862)	*Clarias gariepinus*	F: GGCTAAAAGAACCCTGTCTG R: TACAGCGTCGATAAGCCAGG	54	25	0.1	80-120	5
	Cg05 (U30866)		F: TCCACATTAAGGACAACCACCG R: TTTGCAGTTCACGACTGCCG			0.2	190-230	5
	Cga09 (U30871)		F: CGTCCACTTCCCCTAGAGCG R: CCAGCTGCATTACCATACATGG			0.05	170-210	7
	Cga11 (AF009347)		F: TGCATTTTTGTGTAGTGATGC R:GGGCTCCTGGAAGATTGG			0.05	160-200	5
Singleplex	Cga02 (U30863)	*Clarias gariepinus*	F: GCTAGTGTGAACGCAAGGC R:ACCTCTGAGATAAAACACAGC	54	25	0.2	90-140	5

3.2.3. Genetic diversity

Genetic variation in each population was measured by calculating the mean number of alleles per locus, the observed (H_0) and unbiased expected (H_E) heterozygosities and the $F_{IS.}$ Deviation from Hardy-Weinberg equilibrium and linkage disequilibrium (LD) between pairs of loci (significant with exact test) were calculated at each location, each region and across all samples with GENETIX v. 4.05 (Belkhir *et al*., 2004).

3.2.4. Patterns of population subdivision

We used various methods to describe the population substructure of *C. gariepinus* of the Congo Basin. First, population differentiation was quantified in GENETIX using the standardized allelic variation (F_{ST}) estimated as θ (Weir & Cockerham, 1984) and in SPAGeDI 1.2g (Hardy & Vekemans, 2002) using R_{ST} an analogue of F_{ST} for microsatellites estimated as ϱ (Skaltin, 1995). F_{ST} linked pairwise genetic distance were calculated according to Cavalli-Sforza and Edwards (1967) (D_{CE}) with GENETIX, and R_{ST} linked pairwise genetic distances according to Goldstein *et al*. (1995) ($d\mu^2$) with SPAGeDi. We used 10^4 random permutations of data to test for significance of F_{ST} and D_{CE} in GENETIX. The significance for R_{ST} and $d\mu^2$ were tested in SPAGeDi against 10^4 random permutations.

Second, the distributions of the expected heterozygosity for each locus and population under both the Infinite Allele Model (IAM) and the Stepwise Mutation Model (SMM) were calculated using BOTTLENECK (Cornuet & Luikart, 1996).

Thrid, classical multidimentional scaling analyses (MDSA) based on the genetic distances were obtained using Statistica 9.0. (StatSoft, 2009).

Fourth, Structure 2.0 (Pritchard *et al*. 2000) was used to assign individuals from populations to a pre-determined number of clusters (*K*) based on multi-locus microsatellite data. 10,000 generations of 'burn-in' and 100,000 Markov chain Monte carlo (MCMC) generations were used for each value of K and individuals were assumed to have a mixed ancestry, with correlated allele frequencies among populations. For each *K* of 1-8, 8 runs were perfomed. Estimated log probabilities (Ln P (D)) were average across run and compared to determine the posterior probability of each *K*.

Finally, to analyse the effect of geographical distance on genetic distance, the Mantel test in GENETIX (Belkhir *et al*., 2004) was used, which computes the correlation between distance matrices by means of the permutation procedure (Mantel, 1967). Geographical distances were calculated using the electronic map Google Earth 2010 (www.Google.com).

3.3. Results

3.3.1. Genetic variation within populations

We used allelic diversity and levels of heterozygosity as indicator for within population genetic diversity. Mean number of alleles, observed and expected heterozygosity values for each population are shown in Table 3.3. The mean number of alleles per locus ranged from 4.6-8.2. High values of the mean number of alleles were detected in the Lutshipuka (LUT) (8.2) and Kisangani (KIS) (7.8) samples. Mean expected heterozygosity was relatively uniform among sampling sites with the highest value in the Mwaba (LUT) (0.695) and Kisangani (0.686) locations and lowest in the Kapolowe Station locations. The other locations showed similar moderate values (Table 3.3). All sites (DLB, KIS, KAD, KAF, LUT, KYU and LUF) showed significant departure from HWE ($F_{IS} > 0$) except for KIN (Table 3.3). But, only LUF and KAF showed significant departure from HWE for a 95% C.I. Pair-wise comparisons between loci revealed no significant linkage disequilibrium after sequential Bonferroni corrections.

Table 3.3. Estimates of genetic diversity at eight sites of *Clarias gariepinus* based on five microsatellite markers. $H_{En.b.}$ is the unbiased expected heterozygosity, H_O the observed heterozygosity, and F_{IS} measures the deviation from Hardy-Weinberg equilibrium. See Table 3.1 for sample codes.

Site	Population	Sample size	Mean number of alles per locus	$H_{e.n.b.}$	H_O	F_{IS}
1	DLB	47	6.0	0.574	0.515	0.103
2	KIN	20	4.6	0.538	0.641	-0.196
3	KIS	38	7.8	0.686	0.675	0.016
4	KAD	26	6.0	0.688	0.599	0.132
5	KAF	25	5.8	0.576	0.392	0.061
6	LUT	29	8.2	0.695	0.691	0.006
7	KYU	48	5.8	0.475	0.470	0.058
8	LUF	47	5.0	0.524	0.353	0.328

3.3.2. Patterns of population subdivisions

First, the global standardised multilocus F_{ST} and R_{ST} was respectively 0.290 and 0.442, a value indicating to high level of differentiation in comparison with the results (F_{ST}=0.144 and R_{ST}=0.219) of Wacirachaikarn *et al.* (2009). The pairwise F_{ST} value was significant for all populations (Table 3.4). The Kisangani and Kinshasa populations showed a low value of F_{ST} (0.038). The pairwise R_{ST} value was significant after Bonferroni correction for a limited number of population pairs between Kinshasa (KIN) and Upper Lufira (LUF); Kisangani (KIS) and Kafubu (KAF), Kisangani (KIS) and Lutshipuka (LUT), Kisangani (KIS) and Kyubo (KYU); Kando (KAD) and Lutshipuka; Kando (KAD) and Upper Lufira (LUF).

Second, the classical Multidimensional Scaling plots based on the genetic distances of Goldstein *et al*. (1995) separated clearly the group of Congo-Lualaba (Bukama, Kisangani and Kinshasa) from the Luapula-Mweru cluster, the Upper Lufira cluster and the middle Lufira cluster (Fig 3.2).

Third, the five microsatellite genotypes did not clearly reveal a high degree of population structure among populations. To observe the closer similarly between the Bukama (DLB) and Kinshasa (KIN) populations, than the rest (Fig 3.3). The results of the Bayesian cluster analysis (Figure 3.4.) pointed to a K=8, without reading a maximum of ΔK to identify the optimal number of groups.

Fourth, Mantel tests revealed not significant between the geographic distance and the genetic distances D_{CE} (r= 0.170; p>0.05) and $d\mu^2$ (r= 0.013; p>0.05).

Table 3.4. Pairwise F_{ST} (above diagonal) and R_{ST} (below diagonal) values of *Clarias gariepinus* at eight sites based on five microsatellite markers. Statistical significant values are listed in bold. See Table 3.1 for sample codes.

	DLB	KIN	KIS	KAD	KAF	LUT	KYU	LUF
DLB		**0.108**	**0.085**	**0.239**	**0.401**	**0.311**	**0.303**	**0.420**
KIN	0.125		**0.038**	**0.235**	**0.351**	**0.273**	**0.307**	**0.437**
KIS	0.092	-0.017		**0.176**	**0.281**	**0.213**	**0.243**	**0.366**
KAD	0.489	0.409	0.408		**0.212**	**0.163**	**0.183**	**0.339**
KAF	0.498	0.439	**0.425**	0.389		**0.113**	**0.308**	**0.424**
LUT	0.382	0.418	**0.408**	**0.323**	0.245		**0.203**	**0.352**
KYU	0.489	0.536	**0.516**	0.299	0.257	0.323		**0.198**
LUF	0.705	**0.735**	0.713	**0.549**	0.599	0.292	0.233	

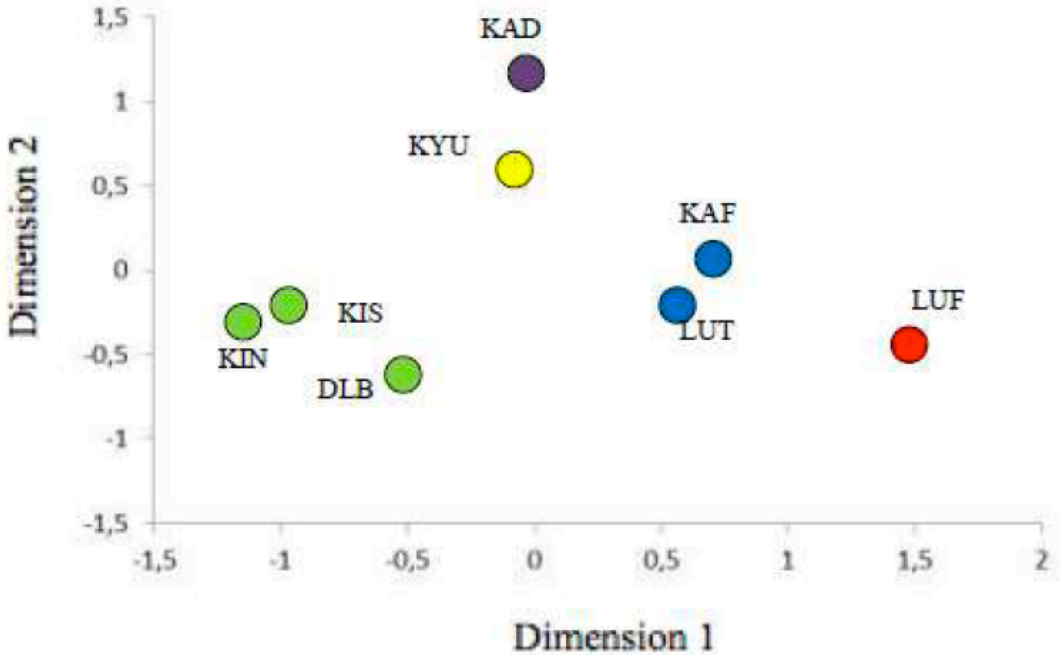


Figure 3.2. Classical multidimensional scaling plots of pairwise genetic distance for the microsatellite data calculated according to Goldstein *et al.* (1995) ($d\mu^2$) of the eight samples of *Clarias gariepinus*. See Table 3.1 for samples codes and Fig 3.1. for the colour of haplotypes.

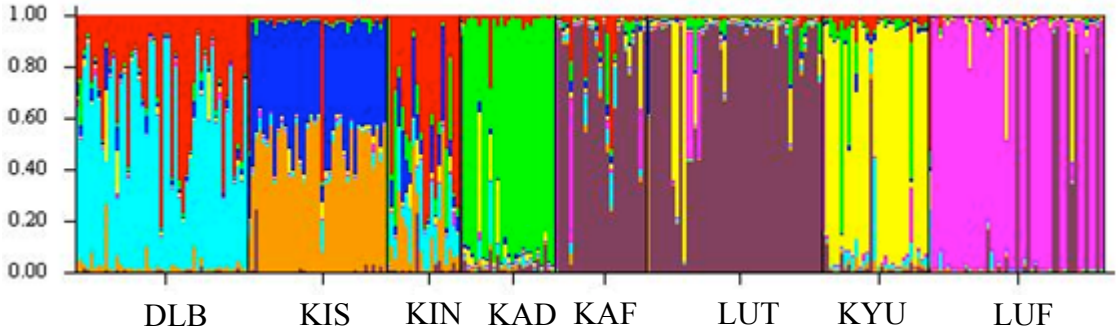


Figure 3.3. Results of the clustering analysis conducted in STRUCTURE 2.2 (Prichard *et al*.2000) based on the microsatellite data. See Table 3.1. for samples codes.

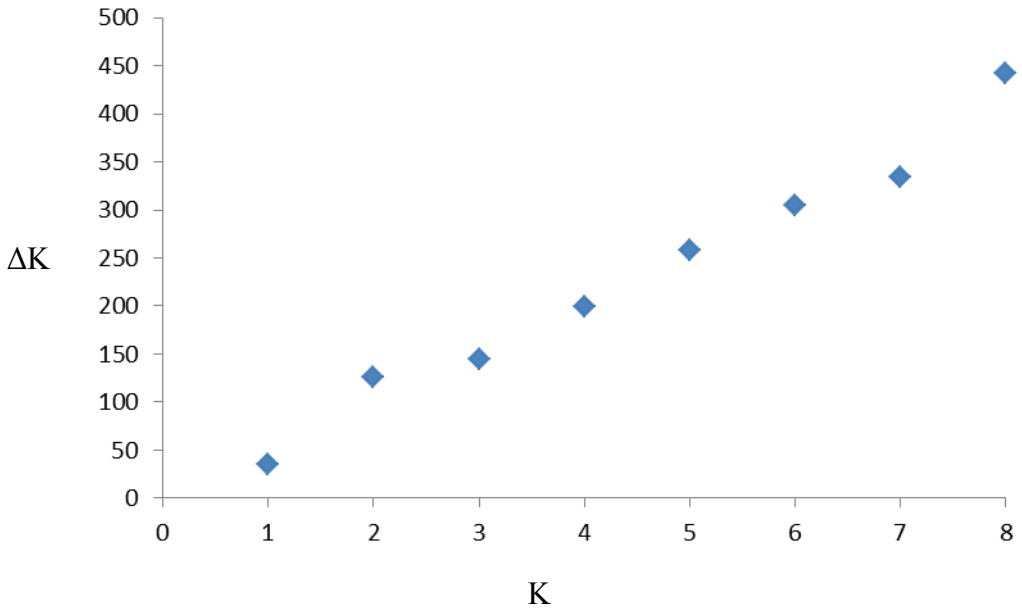


Figure 3.4. Results of the Bayesian Cluster analysis; at K=8 the optimal grouping has not been reached yet.

3.4. Discussion

The main finding of our study is the contemporary isolation of the catfish inhabiting the various drainages of the upper Congo basin in the Katanga province, while fish living in the Congo River show a high degree of relatedness over distances of more than 2500 km. These results agree with the findings from a previous study which focused on a historical perspective that was founded on mitochondrial DNA.

3.4.1. Genetic diversity

On average, heterozygosity values are higher than observed for most freshwater fishes (DeWoody & Avise, 2000). However, heterozygosity levels and the number of alleles of microsatellites may vary considerably from locus to locus. The values observed on *Clarias gariepinus*, except for the samples from Kisangani and Luthsipuka, are lower than recorded by Galbusera (1997) on populations from all across Africa and the Middle East. The lower number of loci included in our data set might provide an explanation (Evanno *et al*., 2005). Also, two of the seven loci are from the conspecific species *Clarias batrachus*. Despite its common genus name, there is growing evidence that Southeast Asian and Africa catfishes do not share a recent common history. Phylogenetically distinct taxa often have either non-conserved primer sites of the microsatellites or invariable loci, and hence a higher chance for monomorphism.

The sample at Kapolowe Station (LUF) showed the lowest heterozygosity and highest level of inbreeding despite the fair number of samples collected (n=47). Also Bukama (DLB), Mwaba (LUT) and Dianda (KAD) showed a significant departure from Hardy-Weinberg equilibrium. At the first sites this may be attributed to bottlenecking following isolation after construction of the Mwadingusha hydroelectric dam. A bottleneck is always accompanied by homozygote excess. Another reason for homozygote excess is the mixing of genetically distinct populations (Wahlund effect; Allendorf & Luikart, 2007), but local evidence seems to be limited judged from mitochondrial haplotype variation. The bottleneck test does not confirm this hypothesis. However, more likely is that pollution from copper and cobalt mining have depressed population sizes, and hence caused genetic drift (Katemo, 2009). Metal pollution has been documented to affect survival with selection for more heterozygous individuals (Roark and Brown, 1996; Maes *et al*., 2005)**.** A large difference between the observed heterozygosity compared to the expected heterozygosity in the Kinshasa sample might reflect outbreeding or mixing of populations. Pool Malebo has the characteristics of a large lake (500 km²) crossed by the Congo River. However, the Kinshasa sample was small (n=20) and hence not very representative.

The number of alleles per locus is on average lower than reported in the literature (DeWoody & Avise, 2000). They are higher for microsatellites isolated from *C.*

gariepinus compared to *C. batrachus*. Galbusera *et al*. (1996) and Wachirachaikarn *et al*. (2009) detected for microsatellites from *C. gariepinus* a number of alleles per locus ranging from 5 to 14 and 4.67 to 12.17 respectively, comparable to our results. However for microsatellites from *C. batrachus,* the number of alleles per locus detected by Yue *et al.* (2003) is lower than reported here.

3.4.2. Population differentiation

Many studies on the population structure of *Clarias gariepinus* have shown the presence of distinct populations regardless of the genetic marker used (Galbusera, 1997; Van der Bank *et al.*, 1992; Van der Walt *et al*., 1993; Giddelo *et al*., 2002; Wachichairkan *et al*., 2009). Our results point to a basin-specific separation in the Katanga region and a high level of gene flow over a wide range of the Congo River. Such results point to old patterns of differentiation in the first case and recent expansions in the second case. This interpretation matches well with the mitochondrial haplotype variation in the Congo Basin as studied by Chocha *et al*. (submitted; Chapter 2.). The distinct mitochondrial differentiation between fish of the Congo-Lualaba cluster, Luapula-Mweru, Upper Lufira and Middle Lufira matches with the assigned microsatellite populations. Ancestral units characterise the upstream regions of the Congo basin. In the mitochondrial Congo-Lualaba cluster occupying the Lualaba and Congo River which slowly drains the *Cuvette Centrale*, differentiation between populations was very low and suggested a recent pattern of gene flow and population expansion. Despite the large distance between upstream (Bukama; DLB) and downstream (Kinshasa; KIN) populations, the genetic differentiation obtained with two different markers was lower than the populations inhabiting the Katanga plateau. The results confirm a local evolution in the upstream (southern) Congo basin and the large dispersal capacity in the Central Congo Basin.

mtDNA has a relatively fast rate of nucleotide divergence, well suited for examining events over the last few million years. For more recent events (last 10,000 years) other markers, such as microsatellites, are more suitable (Hewitt, 2004). In fact in our case the microsatellite markers are suboptimal for large-scale studies of population genetic

structure. If R_{ST} is larger than F_{ST} populations diverged for a sufficiently long time and/or populations exchanged migrants at a rate similar or inferior to the mutation rate (Hardy *et al*., 2003).

3.4.3. Does the biodiversity of catfish in the Congo basin point to a Museum, large-distance dispersal or refugee function?

The biogeographical patterns of the fishes of the large rivers are testimony to numerous and various past events (Thieme *et al*., 2005; Renno *et al.,* 2006). Hypotheses on the nature of the dynamics of accumulation have been qualified under a range of models, including the museum, refuge and long-distance dispersal model. They have provided a suitable reference frame to explain biogeographical patterns in large rivers (Hubert & Renno, 2006). However, to which extent does this hold at the phylogeographical level? For example Hudbert *et al*. (2007) implicitly attribute the museum hypothesis to the patterns of the piranha *Serrasalmus* in the Amazon basin. We discuss here the above mentioned hypotheses in the perspective of the catfish *Clarias gariepinus*.

The Museum hypothesis points to a long and steady accumulation of biodiversity. Unlike temperate and polar biota, tropical biota has never been subject to radical population extinctions. Changes in precipitation and temperature have affected the coverage of forests (Maley, 2001). Unlike several other large tropical large rivers (Mekong and Amazon), no late Tertiary and Quaternary marine incursions characterize the Congo basin. For example, the divergence time between the freshwater herring of lake Tanganyika and marine herring species dates from between 25 and 50 MYA (Wilson *et al*., 2008). Hence no marine taxa are found. The lack of anadromous and highly migratory fish taxa in the Congo river is another remarkable observation linked to the absence of a marine context. For example several marine families (Sciaenidae, Engraulididae and Belonidae) and a cetacean (*Inia geoffrensis)* inhabit the Amazon, and migratory catfishes (Pimelodidae) spawn upstream and grow up downstream. The Mekong river harbours marine fish families of the Gobiidae (Rainboth, 1996), a cetacean (*Orcaella brevirostris*) and large migratory catfishes of

the Pangasiidae family (Hogan *et al*., 2007). Hence, the accumulation of taxa reflects steady local speciation over a long period and supports the museum hypothesis. However, the biogeography of the fish fauna is to poorly documented to support the museum hypothesis. Surely the presence of catfish in the Congo basin proper cannot be attributed to a long steady evolution. It has invaded the area relatively recently.

The faunas at the periphery of the Congo basin enjoyed a seasonal climate over a long geological period and show evidence of faunal accumulation in the Katanga region (Chapter 2; Poll, 1976, Katongo *et al*., 2005, 2007) and the Great Lakes (Lévèque, 1997).

Long-distance dispersal is an important source of gene flow; it refers to the colonization from areas at some distance, although habitat continuity is an important contributor to successful colonization. Moreover, habitats with vacant niches are more likely to be successfully colonized (monopolization hypothesis; Cottenie & De Meester, 2004).). Genetic differentiation is influenced by the dispersal capacity of the taxon considered and may be proportional to geographical distance. Long distance dispersal does seem to be a factor of some significance in large rivers. For example pirarucu *Aparaima gigas* migrates over long distances in the Amazon River without evidence for a correlation between geographical and genetic distance (Hrbek *et al.,* 2005). The catfishes *Pangasius bocourti* and *Pangasianodon hypophthalmus* and the snakehead *Channa striata* show no evidence of isolation by distance in the Mekong river (So *et al*., 2006; Adamson, 2010). *C. gariepinus* shows low nuclear and mitochondrial differentiation between populations caught in Bukama and Kinshasa separated by a distance of 2500 km. It confirms its large dispersal capacity.

The refuge hypothesis predicts that suitable habitats must have been limited because of the Pleistocene climatic fluctuations. Tropical rainforests were fragmented during glacial periods. In the Congo basin reduced precipitation and lower temperatures led to a replacement by tropical seasonal forests (Maley, 2001). The associated forest flora and fauna retreated to a few refuges, while the organisms adapted to the new conditions expanded. The Mayombe refuge in the lower Congo basin is well accepted as one of the four tropical African refuges (Dainou *et al*., 2010). Refuges in the savannah have been established in the south, east and north (Hewitt, 2004). Katongo *et*

al. (2005) and Chocha *et al.* (Chapter 2) added a fourth refuge for the haplochromine cichlids and *C. gariepinus* respectively in the southern Congo basin. The currently used nuclear markers show that the contemporary population of *C. gariepinus* conserved the refuge state, which provides additional support for the refugee hypothesis.

In conclusion the genetic patterns of *C. gariepinus* in the Congo basin are affected by a combination of processes such as barriers, a long history of divergence, with an impact of refuges and new dispersal opportunities.

Acknowledgements

We thank J.-P. Marquet (PRODEPAAK/BTCCTB) for help with field sampling. D. Schaerlaekens helped with the scoring of the genotypes, J. Raeymaekers and L. Mbaya helped with the GIS. Funding was provided to ACM by the Belgian Technical Cooperation, the K.U.Leuven to MHDL and the Research Foundation-Flanders (FWO-Vlaanderen) to MPMV.

B. Hellemans helped with molecular biology.

Chapter 4

Phylogeographical patterns of the Congo Basin: understanding of the Pan-African evolution of the catfish *Clarias gariepinus*

Auguste Chocha Manda, Maarten Vanhove, Maarten Larmuseau, Karine Van Doninck, Jean Claude Micha, Filip A.M. Volckaert

Manuscript in preparation

Abstract

During the Miocene, faunal barriers between the hydrographical systems of Africa were less distinct than today. Climatic fluctuations and geological events changed the morphology of the rivers and ancient connections between the various basins. The phylogeographical pattern of the catfish *Clarias gariepinus* was studied in the Congo Basin and in the neighboring basins of Nile, Zambezi and Eastern Africa to confirm the scenario of ancient contacts across the ichthyological provinces during the Pleistocene. In total 133 *cytochrome b* sequences of 509 bp were aligned and their patterns analysed. Thirty seven unique haplotypes were identified and clustered phylogenetically in four clades: Nilo-Sudan, Zambezi, East-Coast and Congo. Network analysis confirmed the results and showed the high diversity of the Congo clade. The ancestral haplotypes of three Congo clades extended into neighboring basins. It confirms that fish genomes are witnesses of the past and that their evolution harbours much information on past global changes.

4.1. Introduction

During the Miocene, some 25 million years ago, faunal barriers between the hydrographical systems of Africa were less distinct than today. A rather uniform and widespread fish fauna inhabited tropical and subtropical Africa (Lévêque, 1997; Roberts, 1975; Beadle, 1981). During the later Tertiary and Pleistocene climate and geology (tectonics and erosion) shaped the local habitats (Maley, 2001) and fauna (deMenocal, 2004). Communities characteristic of the various habitats developed and matured. The African continent was subjected to strong climatic fluctuations with the succession of dry and wet phases, which have profoundly acted on the hydrographical systems and consequently, on the geographical distribution of fish species (Roberts, 1975). These elements contributed to the delineation of the various ichthyological provinces in Africa as proposed by Roberts (1975) and later reformulated by Lévêque (1997).

The affinity between the fish faunas of the various ichthyological provinces in Central Africa has been reported in literature by Poll (1963, 1976), Lévêque (1997), Skelton (1994) and Thieme *et al.* (2005). The Congo basin represents the largest and richest ichthyoprovince of Africa (Thieme *et al.*, 2005). It borders the Nilo-Sudanic province in the north, the East Coast in the east, the Zambezi in the south and the Lower Guinea Atlantic (tropical Africa province) with whom it shares many affinities in the west (Lévêque, 1997). The vicinity of other fish faunas and the varying connectivity due to changes in river flow and river capture makes that elements of neighboring basins have been reported in the Congo basin, including the Zambezian cichlids *Oreochromis macrochir* and *Serranochromis sp.* (Katongo *et al*., 2007), the catfish *Clarias ngamensis* (Poll, 1976) and the Nile perch *Lates niloticus* (Lévêque, 1997).

Several scenarios have been proposed to explain the presence of Nilotic and Zambezian fish elements in the Congo basin. The exchange of fishes across tropical northern Congo and the subtropical watershed of the Chad basin via the upper reaches of the Ubangi River have been attributed to river capture (Lévêque, 1997). To Poll (1963) the presence of Nilotic fishes in the Upper Lualaba is evidence of an ancestral link with the Nile at Porte de l'Enfer. The Zambezian fish fauna was probably introduced in the Congo basin under the impact of the superswell of southern Africa

(Thieme *et al.,* 2005; Lévêque, 1997; Poll, 1976). The continental rise in altitude and the tipping of the local topography led to many changes in flow direction (Skelton, 1994). Bell-Cross (1965) suggested that the Upper Zambezi/Okavango and Kafue systems formed part of the late Tertiary western drainage basin whose fauna showed affinities with the Congo-River system (rivers Kasai and Lualaba). Joyce *et al*. (2005) noticed numerous sympatric haplochromines of diverse shape and size in the upper Congo and middle/upper Zambezi, as far south as the Okavango, Cunene and Limpopo rivers. Agnèse & Teugels (2005) who studied the *Clarias gariepinus* and *Bathyclarias* phylogeny and Katongo *et al*. (2007) who studied the serranochromine cichlids corroborated these scenarios.

Based on extensive phylogeographical studies of mostly ungulates (e.g. Nersting and Arctander, 2001; Lorenzen *et al*., 2006; Brown *et al*., 2007) Hewitt (2004) proposed that their expansion was carried by colonization of the savannah during the Pleistocene. The period matches with the separation of most ichthyological provinces (Lévêque, 1997). However a sound proof of this event is missing as because unlike large mammals few fish species have a wide distribution across Africa. An exception is the catfish *Clarias gariepinus*, which lives in a wide range of ichthyological provinces across Africa, including dry and wet savannah, and (sub)tropical forest up to 1600 m altitude (Teugels, 1986).

Here, we propose to study the phylogeographical pattern of the catfish *Clarias gariepinus* in the Congo Basin and in the neighboring basins of Nile, Zambezi and East Coast to confirm the scenario of ancient contacts across the ichthyological provinces during the Pleistocene.

4.2. Materials and methods

4.2.1. Sampling and mtDNA sequencing

We used a total of 15 adult *C. gariepinus* from the Zambezi and Nile basins and Lake Victoria, Tana River (Fig. 1, Table 1). These samples were collected and preserved in pure ethanol by LADS. Genomic DNA was extracted from barbell clips using a NucleoSpin®96 Tissue Kit (Macherey-Nagel GmbH Co.KG) following the

manufacturer's instructions. The primers used, L152 67, 5'-AAT GAC TTG AAG AAC CAC CGT-3' and H15891, 5'-GTT TGA TCC CGT TTC GTG TA-3' (Briolay et *al*., 1998), amplified a fragment of 660 bp of the mitochondrial cytochrome *b* (cyt *b*). PCR reactions were performed in a volume of 25 µl containing 2.5 µl of 10xPCR buffer (Eurogentec), 1 µl 50 mM $MgCl_2$ (Eurogentec), 2.5 µl of 2 mM dNTPs (Amersham Pharmacia Biotech), 1 µl of each primer (20 µM), 0.1 µl of 5 U/µl Silverstar Taq polymerase (Eurogentec) and 16 µl of milli-Q H_2O. Cycling conditions were as follows: initial denaturation at 95°C for 3 min, 35 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 45 s, and a final extension step of 72° for 7 min. After purification from the agarose gel with the GFX Purification Kit (Amersham Pharmacia Biotech), sequencing was done using the same primers as above, with the Big Dye Terminator 3.1 kit (Applied Biosystems), applying a 1/8 dilution of the Big Dye Terminator sequencing protocol. Products were finally run on an ABI 3130 Genetic Analyser (Applied Biosystems). Sequence data were analyzed using the sequencing analysis software SeqScape v.2.5 (Applied Biosystems). Sequences were deposited in GenBank under accession numbers XXX-XXX *(numbers available after revision)*.

4.2.2. Alignment and phylogenetic analysis

A total of 15 cyt *b* sequences generated during this study were aligned together with 118 sequences (except from lake Kivu) from the study of Chocha Manda *et al*. (submitted) and published GenBank sequences of *C. gariepinus* from Lake Mazalah in Egypt (AF235924) and lake Malawi in Malawi (AF235922) (Agnèse & Teugels, 2005), Gangi River in Nigeria (GU906881, GU906880 and GU906879) and Lake Victoria in Kenya (DQ646360, DQ646361, DQ646362, DQ646365, DQ646370, DQ646371, DQ646372) (Mwita and Nkwengulila, 2008). The alignment was realized using the program Muscle 3.6 (Edgar, 2004) and further improved by TrimAL 1.2 (Capella-Gutiérrez *et al.*, 2009). Three haplotypes of *Clarias ngamensis* (AF23593, XXX and XXX) were selected as outgroup for the phylogenetic analyses. The appropriate evolutionary model was selected with jModelTest

(Posada, 2008). According to the Akaike information criterion (AIC), the TPM2uf (Posada, 2009) + Γ model was the most suitable model for further analysis, with a gamma shape parameter of 0.18. Two complementary algorithms were used, maximum likelihood (ML) and Bayesian inference (BI) implemented respectively in PhyML 3.0 (Guindon & Gascuel, 2003; Guindon *et al*., 2009) and MrBayes v. 3.1.2 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). For ML, the robustness of the nodes was assessed by bootstrap analysis with 1000 replicates. The general time reversible (GTR; Tavaré, 1986) model with a gamma distribution of site-specific rates was used in the Bayesian analysis because it is the next most complex model to TPM2uf + Γ available in MrBayes v 3.1.2. Posterior probabilities were calculated over $2.5x10^6$ generations, while sampling the Markov Chain at a frequency of 100 generations. 25 % of the samples were discarded as "burn-in''.

4.2.3. Network analysis and phylogeographical history

A haplotypes network was constructed with the software TCS v. 1.3, as described by Clement *et al.* (2000) and NETWORK v. 4.5.0.1 (http://www.fluxus-engineering.com). Using the program DnaSP v. 5 (Librado & Rozas, 2009), cyt *b* polymorphism was estimated as nucleotide (π: Nei, 1987) and haplotype diversity (h_d: Nei & Tajima, 1981) for each observed clade.

Each sample and drainage (pooled sample) was tested with three classes of methods based on the information they incorporate (Ramos-Onsins & Rozas, 2002) for deviation from mutation-drift and gene flow-drift equilibrium. Deviation from neutral expectations may be evidence for past demographic events such as population growth or the presence of population sub-structuring within a sample. Tajima's *D* (Tajima, 1989) that compares the number of segregating sites to nucleotide diversity in a sample (Class I) was used to test for deviation from neutrality due to selection, population bottleneck, or admixture (Rand, 1996). Fu's F_S (Fu, 1997) that compares θ estimated from nucleotide diversity with the expected number of haplotypes Ewens(1972) distribution given the sample size (Class II) was used to detect past fluctuation in population size, and R_2 (Ramos-Onsins & Rozas, 2002) that uses

information from the mismatch distribution of pair-wise differences (Class III) was also employed to examine demographic events. Tajima's D, Fu's F_S and Rozas' R_2 were calculated in DnaSP v. 5 (Librado & Rozas, 2009). Significance values for all three tests were calculated using coalescent simulations implemented in DNASP (with 1000 replicates for each simulation).

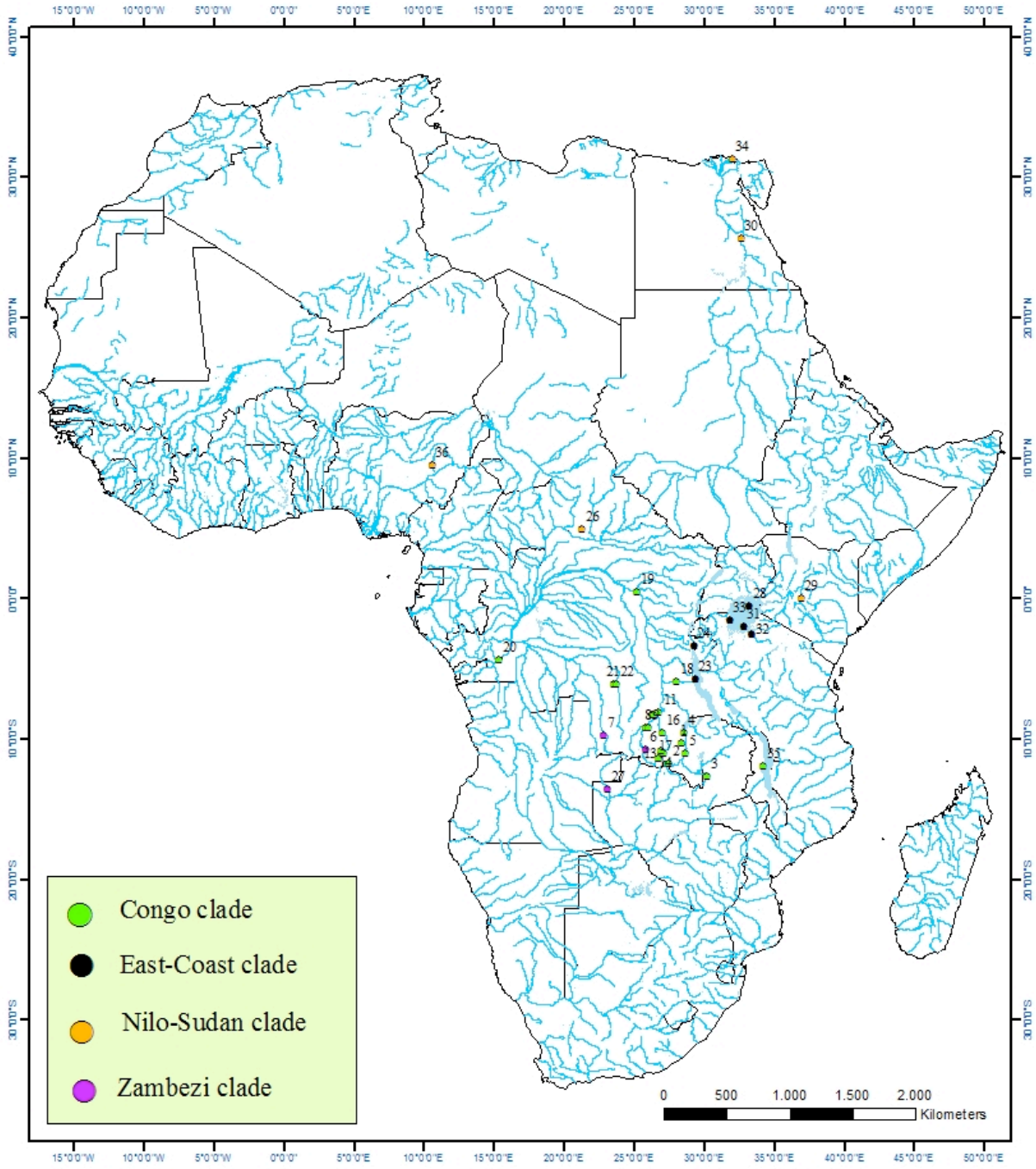


Figure 4.1. Geographical origin of the samples of *C. gariepinus* (for numbers, see Table 2.1 and Table 4.1)

Table 4.1. Geographical origin of the samples of *Clarias gariepinus*, number of samples examined (N) and unique cytochrome *b* haplotypes. R, River

Site	Code	River or Lake	Location	Geographic coordinates		N	Haplotypes
27	ZZZ	Zambezi R	-	-	-	5	CGA004,CGA007,CGA038, CGA39, CGA040
28	VIC	Lake Victoria	-	-	-	3	CGA20,CGA028, CGA029
29	TAN	Tana River	-	-	-	4	CGA030,CGA031,CGA032,CGA033,
30	LOX	Nile River	Luxor	-	-	3	CGA035,CGA036,CGA037
31	UKE	Lake Victoria	Ukerewe island	-	-	2	CGA020
32	MAB	Lake Victoria	Magu Bay	-	-	2	CGA027
33	KGR	Kagera River	-	-	-	3	CGA028
34	MAN	Lake Manzalah	-	-	-	1	CGA037
35	MLW	Lake Malawi	-	-	-	1	CGA025
36	GAR	Gaji River	-	9°28'N	10°34'W	3	CGA034

4.3. Results

4.3.1. mtDNA haplotypes

The alignment of sequences was straightforward as there were no gaps and translation into amino acids did not indicate nonsense or stop codons. The sequence characteristics matched the general properties of the *C. gariepinus* cyt *b* gene (Agnèse & Teugels, 2001a), suggesting a functional mtDNA cyt *b* gene and not a nuclear pseudogene (Zhang & Hewitt, 1996).

After alignment of the 146 sequences, 37 haplotypes were recognized. Our dataset of 509 bp contained a total of 65 variable sites, of which 49 were parsimony informative.

4.3.2. Phylogenetic analyses

Phylogenetic analysis of the 37 unique sequences revealed a Nilo-Sudan clade, Congo clade, East-Coast clade and Zambezi clade well-supported by bootstrap and posterior probability values (Fig. 4.2). The topologies of ML and BI trees were almost identical, with only small differences. The Congo clade united 16 haplotypes from the rivers Congo, Lualaba, Lufira and from the Luapula-Mweru system. The haplotype of Lake Malawi is included in the Congo Basin clade. The Zambezi clade shows six haplotypes from the Lulua, Kando and Zambezi rivers. The East-Coast clade shows four uniques haplotypes from the Lake Tanganyika system, the Victoria system and the Rusizi River. The Nilo-Sudan clade shows nine haplotypes from Lake Mazalah, Lake Tana, the Ubangi system and the Gagi River (Fig. 4.2).

Table 4.2. Genetic diversity of the populations of *C. gariepinus* based on mitochondrial cytochrome *b* data: number of specimens (N), number of haplotypes (N_H), number of polymorphic sites (P_S), haplotype diversity (h_d) and nucleotide diversity (π)

Unit	N	N_H	P_S	h_d (S.D.)	π (S.D.)
Total data	146	37	65	0.870 (0.00037)	0.01968 (0.02378)
Congo clade	102	16	35	0.761 (0.00112)	0.00959 (0.00959)
East-Cost clade	19	5	7	0.386 (0.01930)	0.00182 (0.00395)
Zambezi clade	13	7	19	0.795 (0.01191)	0.00298 (0.00445)
Nilo-Sudan clade	12	9	7	0.939 (0.00333)	0.01614 (0.01241)

Table 4.3. Demographic analyses of *C. gariepinus,* including Tajima's *D*, Fu's F_S test and R_2 test

	Tajima' s *D*	Fu' s F_S	R_2
Congo clade	-0.10644	-0.28906	0.09201**
East-Coast clade	-0.02315	0.19469	0.16344**
Zambezi clade	-0.07261	0.16085	0.17679**
Nilo-Sudan clade	-0.11926	0.18655	0.15267**

**p<0.01

4.3.3. Network and demographic analyses

Phylogeographic reconstruction using median-joining parsimony and TCS 1.2 showed four different clades: the Congo clade, the East-Coast clade, the Zambezi clade and the Nilo-Sudan clade (Fig. 4.2). With TCS 1.2, the Congo clade shows Cga008 as an ancestral haplotype which is found in the Congo-Lualaba system. The East-Coast clade shows Cga020, which is found at all locations. The Zambezi clade displays Cga004 as an ancestral haplotype, found in the River Kando and the Zambezi system. The genetic diversity (Table 4.2) was higher for the Nilo-Sudan clade (h_d = 0.939 and π = 0.016) than for the Zambezi clade (h_d = 0.795 and π = 0.002), Congo clade (h_d = 0.761 and π = 0.009) and East-Coast clade (h_d = 0.386 and π = 0.001). Table 4.3 summarizes the results of the clade specific demographic history.

Table 4.3 summarizes the results of the group-specific demographic history. Possible evidence for a recent range expansion of all clades (Congo, East Cost, Zambezi and Nilo-Sudan clades) supported by significant values for Ramos-Onsins and Rozas' R_2 respectively of R_2=0.09201, R_2=0.16344, R_2=0.17679, R_2=0.15267).

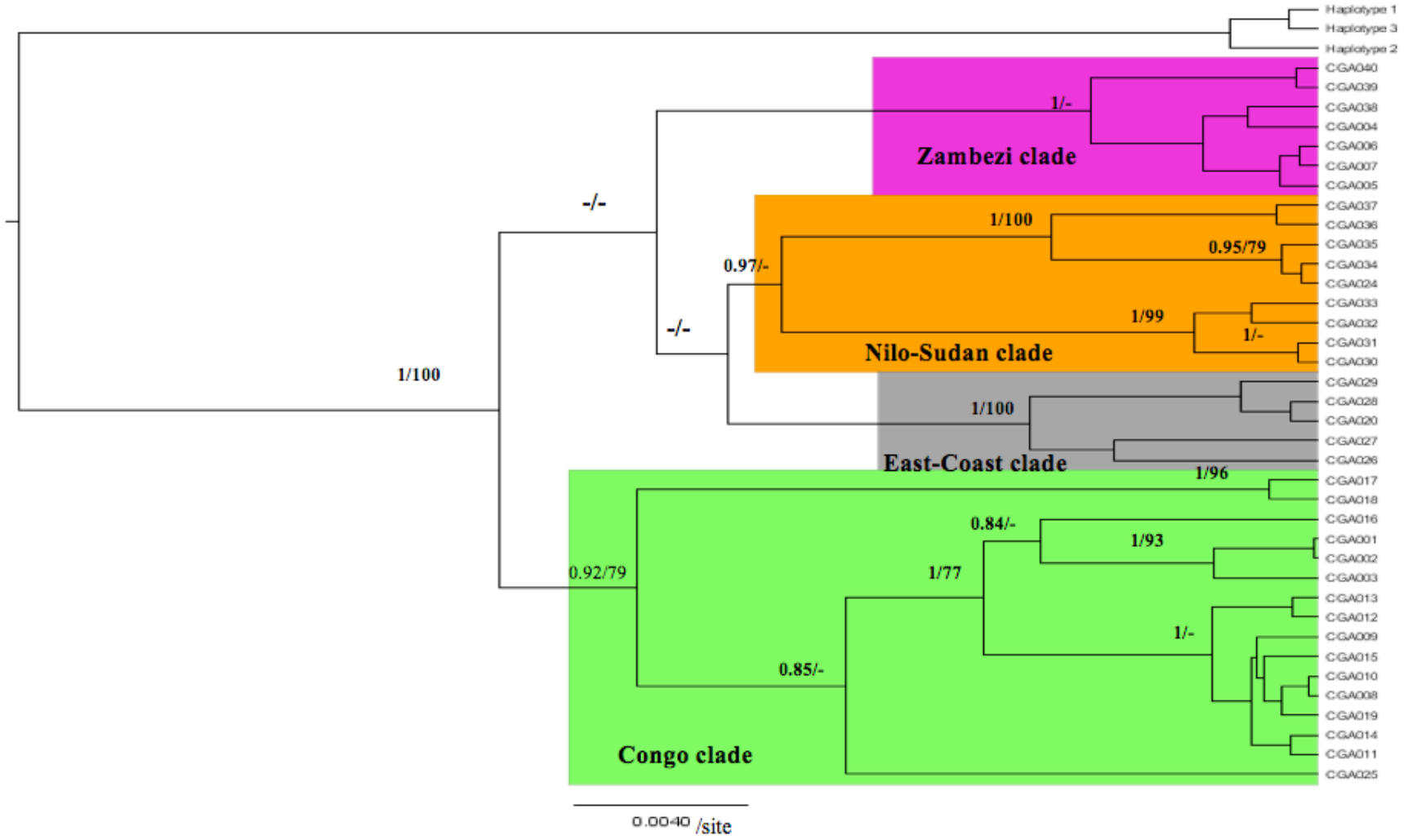


Figure 4.2. Bayesian phylogenetic tree of *C. gariepinus* populations in Africa. The tree was reconstructed based on the *cytb* haplotypes found in 146 individuals of *C. gariepinus*. Numbers beside nodes are from left to right: Bayesian posterior probabilities and Maximum likelihood bootstrap support. Values below 0.80 (Bayesian posterior probabilities) or 70% (Maximum likelihood bootstrap support) are marked with '-'.

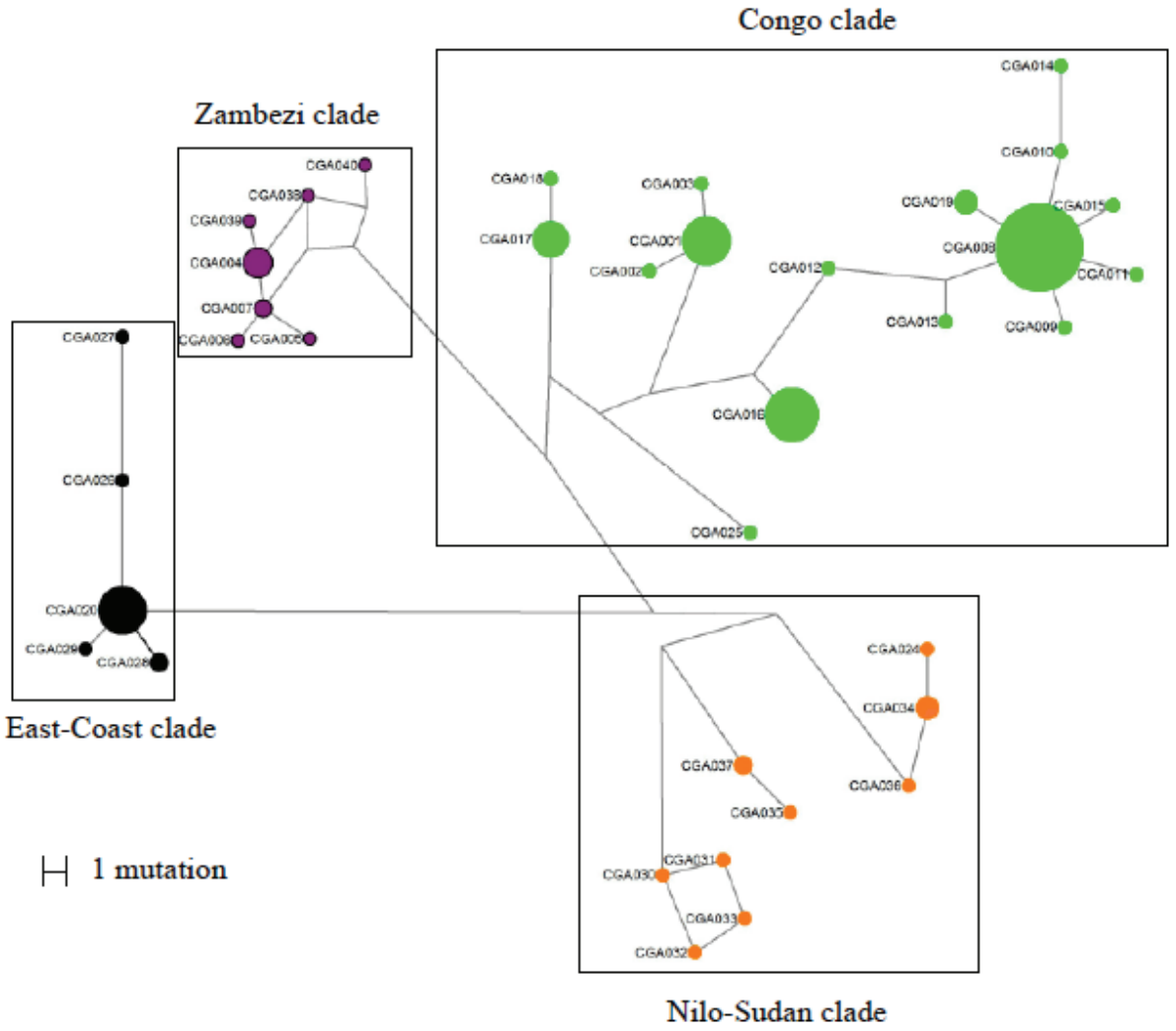


Figure 4.3. Median-joining network of mitochondrial DNA cytochrome *b* haplotype in African *C. gariepinus* populations. The size of the circles is proportional to the number of Catfish individuals sharing that haplotype. See Fig 4.1 for the colour of haplotypes.

4.4. Discussion

It is clear from the cyt *b* haplotypes that the core region of the Congo basin is inhabited by an ancestral indigenous clade (Congo clade), which represents one of the four major clades characterising the catfish *Clarias gariepinus*. The catfish inhabiting the peripheral zones of the Congo basin are affiliated with the Nilo-Sudan, East-Coast and Zambezi clades.

4.4.1. Diversity and lineages

The Nilo-Sudan clade shows the highest diversity among all clades. Fragmentation within Northern and Western Africa is consistent with wet/dry cycling in the Sahara region and the concomitant isolation of groups during the Pleistocene (Rognon *et al*., 1998; Agnèse & Teugels, 2001b, 2005; Giddelo *et al*., 2002, Arndt *et al*., 2003; Rognon & Guyomard, 2003; deMenocal, 2009). Diversity of the Central Congo and Zambezi clades is also high. The former results match with Chocha *et al.* (Chapter 2). Both clades have experienced a long history of climate change and hydrological modifications due to tectonic and geological influences. The low diversity of the Eastern Africa clade represents more likely a sampling artifact as Giddelo *et al.* (2002) noticed a highly fragmented environment (see further).

The Nilo-Sudanic clade includes catfish populations of the Nile River in Northern Africa, the Gaji, Niger and Senegal River in Western Africa and the Northern part of the Congo basin including the Ubangi River, and the Lower Tana River in Eastern Africa. The Nilo-Sudanian fish fauna comprise the fishes of the Nile, Chad, Niger, Volta and Senegal (Roberts, 1975; Lévêque, 1997). Although the Ubangi River is part of the northern Congo basin, the fauna of Northern Congo has affiliations with the Nilo-Sudanic fauna (Bailey, 1986). The Tana River flows eastward in the Indian Ocean (Thieme *et al*, 2005) and is currently not connected to the Nile system. Giddelo *et al.* (2002) noticed a strong genetic differentiation between populations in the Upper and Lower Tana which relates to a split between the North African and East African clade. Many studies related the affinities of the Western African fish fauna with the

Nile (Thieme *et al*, 2005). Rognon *et al.* (1998) measured a strong genetic divergence between populations of *C. gariepinus* from the whole Nilo-Sudan system and Lake Victoria (Eastern clade). They estimated divergence at 3.9 MYA, which matches with the formation of the rift lakes. We confirm their conclusion.

Giddelo *et al.* (2002) studied the phylogeography of *C. gariepinus* in Eastern Africa and identified an East and Western Rift group in the East Africa clade, an obvious consequence of rifting in the region. The populations of the Rusizi and Rugo rivers (Tanganyika system) belong to the Western Rift together with Lake Victoria.

The populations of the Lulua river (tributary of the Kasai river) and Kando river (tributary of the Lualaba river) in the Congo basin belong to the same lineage as the Upper Zambezi system (Zambezi clade). Already Bell-Cross (1965a) and Poll (1976) mentioned that the fish fauna of the Upper Zambezi showed affinities with the fish fauna of both rivers.

Chocha *et al*. (Chapter 2) revealed in the Congo basin the presence of two catfish clades, the Lulua-Kando and the Congo-Tanganyika clade. With the few data available for this study, it was not possible to obtain certainty on the status of other clades in the Congo basin. However, this study confirms the presence of the above mentioned clades, with the Lulua-Kando clade corresponding with the Zambezi clade. The Malawi sample, consisting of just one individual analyzed by Agnèse & Teugels (2001b), placed the Malawi system in the Congo clade. Modern lake Malawi flows in the Lower Zambezi through the Shire River but has a fauna more affiliated to Central Africa (Lévêque, 1997). The *Bathyclarias gigas* catfishes living in Lake Malawi share a common ancestor with *C. gariepinus* of the Luapula-Mweru system (Congo basin) and are hence paraphyletic (Agnèse & Teugels, 2001b). They bear testimony to an ancient contact between Lake Malawi and the Congo basin through the Luapula-Mweru system.

4.4.2. A scenario for an ancient connection between the Congo and other ichthyological provinces

During the late Tertiary and Pleistocene hydrological changes have influenced the distribution of species in the African tropics (Gasse, 2000; Skelton, 1994). One important aspect of species redistribution are river captures, which have played a significant role especially in the southern part of the Congo basin. Former connections between the Zambezi and Congo basin have been documented by many (Poll, 1963, 1976; Bell-Cross, 1965; Lévêque, 1997; Thieme *et al*., 2005). Historically the Luapula-Mweru system has been part of the Zambezi system during the early Tertiary (Skelton, 1994) and showed a link between the Upper Zambezi and Congo river until fairly recently. Stankiewichz *et al*. (2004) related that during the Pleistocene the Chambeshi river was captured, first by the Kafue (Zambezi system) and then by Luapula (Congo system).The affinity of cichlids (haplocromines and *Serranochromis)* between the Upper Zambezi and Congo basin provides evidence for the ancient contact (Katongo *et al.* 2005; Joyce *et al.* 2005; Salzburger *et al*., 2005). Many fish of the Katanga plateau show a high affinity with the upper Zambezian fish fauna (Poll, 1976). Also Bell-Cross (1965) suggested that the Upper Zambezi/Okavango and Kafue systems formed part of the late Tertiary western drainage basin whose fish fauna showed affinities with the Congo-River system (the Kasai and Lualaba). The fact that we found a common ancestor in both systems revealed that contact is recent between the Kando and Upper Zambezi rivers. Both rivers are situated on the border of the Congo-Zambezi watershed.

Expansion of the haplochromines towards southern Africa originated in the central Congo River, from which the genus *Pseudocrenilabrus* and some Serranochromine cichlids (Greenwood, 1993) colonized the Upper Congo-Luapula system, and entered the Zambezi system further south via the Zambezi-Congo watershed (Katongo *et al*., 2005; Joyce *et al.,* 2005).

The fish fauna of the Malagarasi river (entering Lake Tanganyika on the eastern shore) and the Lukuga river (draining Lake Tanganyika at the western shore) show a high affinity (Devos, 1991). This provides evidence that the Malagarasi system was connected to the Lukuga river before rifting started. Giddelo *et al*. (2002) observed

that the fish populations of the Western Rift are linked to the very dynamic geology of the region between the Lakes Albert, Victoria and Tanganyika, where the presence of similar haplotypes points to connectivity.

Recent contact between the Congo and Chad basin is less clear, but the exchange of fish species between the northern Congo basin and the Chad basin via the upper Ubangi River has been reported (Lévêque, 1997).

4.4.3. Dispersal routes for colonization within the Congo basin

Fishes with a wide distribution across Africa such as *Clarias gariepinus, Lates niloticus, Hydrocinus vitatus, Heterobranchus longifilis, Shilbe intermedius* and *Auchenoglanis occidentalis*, which live from the Nile up to West Africa, and south at least to the Congo basin are probably relics of a widespread ancestral fauna. It existed already from the Chad basin during the Pleistocene before the isolation of the Congo basin (and possibly the Nile basin) (Lévêque, 1997). *C. gariepinus* probably inhabited the Congo basin before its isolation from other neighbouring basins. In fact, the remains of fossil indicate that the presence of *C. gariepinus* is related at Plesitocene (Léveque, 1997; Stewart, 2001; Lévèque & Paugy, 2006). The presence of different local lineages confirms this hypothesis. The partial morphometric overlapping and genetic clustering of *C. gariepinus* populations of the Swaziland and Orange River, and the Lake Victoria population could suggest that they descended from a common ancestor, different from the Nilo-Sudanian (Rognon *et al*., 1998) or this common ancestral of Lake Victoria populations is the same as the Tanganyika system.

4.4.4. Comparison of evolution patterns between ungulates and Clarias gariepinus

Extensive studies of several larger African mammals provide some interesting insights into how Pleistocene climatic changes modified their range and hence their genetic structure and divergence (Hewitt, 2004; Maley, 2001). Phylogenetic structure of several large mammals show three common clades (West, East and South) (Hewitt, 2004). These clades were influenced by expansion of the savannah and the fauna from major refuges during the Pleistocene. The dispersal route was probably on the West-

East and East-South axis (Brown *et al*., 2007; Lorenzen *et al*., 2006; Nersting & Arctander, 2001; Hewitt, 2004). The evolution of *C. gariepinus* in Africa shows the same pattern (North-West, East-South). The Congo was probably the link between these three major clades in the past.

In conclusion, fish in the Congo basin with a wide African distribution are probably relics and have inhabited the Congo basin before the isolation between different basins in Africa.

Acknowledgements

We thank J.-P. Marquet (PRODEPAAK/BTCCTB) for help with field sampling. D. Schaerlaekens helped with the scoring of the genotypes, J. Raeymaekers and L. Mbaya helped with the GIS. Funding was provided to ACM by the Belgian Technical Cooperation, the K.U.Leuven to MHDL and the Research Foundation-Flanders (FWO-Vlaanderen) to MPMV.

B. Hellemans helped with molecular biology.

Chapter 5

Conclusions et perspectives

5.1 Discussion générale

Le but de cette thèse était d'étudier l'évolution génétique de la faune ichthyologique du bassin du Congo pour un intérêt aquacole et de conservation. Le modèle retenu était celui du poisson chat *Clarias gariepinus* suite à sa large distribution panafricaine. Ce poisson est signalé dans le bassin du Congo de la source à l'embouchure. Pour y arriver, plusieurs échantillons de tissus des poissons chats sauvages ont été collectés à travers le bassin du Congo et ailleurs en Afrique. Grâce à la comparaison de fragments d'ADN mitochondrial obtenus avec le gène *cytochrome b* et l'estimation du temps de divergence (Chapitre 2), des modèles d'évolution ont pu être établis. Ainsi, de ces modèles ont peu retenir : (1) L'existence de plusieurs groupes et probablement de plusieurs clades de populations de l'Afrique isolées depuis longtemps dans le sud contrairement à la Cuvette Centrale du Congo (entre Kisangani et Kinshasa). En effet, très peu d'haplotypes partagés sur une grande surface ont été observés dans la région de la cuvette centrale. Ainsi, ces résultats ont signalé une ancienne évolution locale dans le sud et une expansion récente dans la cuvette centrale (star-like pattern) au cours du Pléistocène. (2) Afin d'évaluer l'impact de ce modèle historique sur l'évolution des populations contemporaines, une étude sur la structure génétique grâce à des marqueurs nucléaires (les microsatellites) a été initiée (Chapitre 3). Les résultats obtenus ont révélé aussi l'isolation des populations contemporaines dans les régions séparées par des barrières (cas du plateau Katangais). Malgré la grande distance qui sépare les populations de Bukama et de Kinshasa, les deux marqueurs utilisés ont signalé une faible différenciation génétique entre ces populations. Certaines hypothèses ont pu être testées pour évaluer le mode d'accumulation de cette diversité dans le bassin du Congo. Sur les trois hypothèses de départ, deux ont été retenues, il s'agit de l'hypothèse sur la longue distance de dispersion dans la cuvette centrale et celle des refuges dans le sud créés par des barrières et les fluctuations du climat du passé. En plus les populations avoisinantes les régions minières ont présenté des

faibles taux d'hétérozygotie. C'est le cas des populations de Kapolowe Gare. Cette étude a pu aussi montrer les limites de l'utilisation des microsatellites dans le cas des populations isolées pendant longtemps. (3) Pour clarifier l'origine des populations reliques dans le bassin du Congo, une étude complémentaire d'évolution régionale incluant les échantillons d'autres provinces ichthyologiques, a été initiée (Chapitre 4). Ainsi grâce au fragment de l'ADN mitochondrial obtenu avec le *cytochrome b*, les résultats ont révélé l'existence de quatre clades dans le bassin du Congo (Nilo-Soudan, Côte Est, Zambèze et Congo). Pour certaines populations du bassin actuel du Congo, les haplotypes ancestraux sont signalés dans les provinces ichthyologiques voisines (Zambèze et Côte Est), cas de Cga004 rencontré dans la rivière Zambèze (province ichthyologique du Zambèze) et la rivière Kando (province ichthyologique du Congo) ce qui confirme le scénario de l'ancienne connectivité entre les différents bassins en Afrique et une plus récente entre le Congo et le Zambèze. C'est le cas de l'ancienne connexion entre le Congo et le Malawi par le système Luapula-Mweru. Les résultats démographiques révèlent que l'évolution de *C. gariepinus* dans les différents bassins est relativement récente et date du Pléistocène. En conclusion cette étude confirme l'état de relique de *C. gariepinus* dans le bassin du Congo. Les ancêtres ont probablement survécu après l'isolation des anciens bassins ; la route qu'aurait empruntée ce poisson pour coloniser ce bassin est probablement d'Est vers le Sud. Pour une gestion durable de la faune et de la flore du bassin du Congo, il existe des préalables dont le plus important à court terme est la réalisation des études biogéographiques et phylogéographiques de la faune et flore de ce bassin. Ainsi, dans les paragraphes ci-dessous nous discuterons sur l'évolution du poisson chat *C. gariepinus*, sur les études phylogéographiques dans le bassin du Congo, sur l'implication de ces études en aquaculture et sur la conservation.

5.1.1. Evolution du poisson chat Clarias gariepinus

Clarias gariepinus est un poisson panafricain (Teugels, 1986) qui a été signalé en plus au Moyen Orient. Son évolution en Afrique semble être liée à sa capacité à vivre dans des milieux bien différents dont certains sont sujets à des assèchements saisonniers (De Graaf & Janssen, 1996; Teugels, 1986) et à l'impact lié aux changements

climatiques et géologiques du passé. En Afrique du Nord et de l'Ouest, et dans le bassin Nilo-Soudanique, ce poisson présente une évolution fragmentée liée à la fluctuation des périodes sèches et humides du Pléistocène (Arndt *et al*., 2003). En effet, cette région a connu des périodes de succession d'expansion et de régression des systèmes aquatiques accompagnée de phénomènes de colonisation ou d'extinction des populations des poissons (Lévèque & Paugy, 2006) lors des périodes arides du Pléistocène. On retrouve deux grands groupes, celui du Nord et celui de l'Ouest. En effet, les populations de l'Oubangui et de la rivière Gaji en Afrique de l'Ouest forment un groupe, tandis que les populations du lac Manzalah au nord de l'Egypte et Louxor sur le Nil représentent un deuxième groupe. L'affinité entre les poissons du Nord et de l'Ouest de l'Afrique est signalée dans la littérature (Lévèque, 1997). On retrouve ce type de modèle d'évolution fragmentée aussi chez les Cichlidae (Rognon & Guyomard, 2003) de la même région. Les études sur la phylogéographie des Cichlidae et des poissons Killi du sous-genre *Chromaphyosemion* en Afrique de l'Ouest révèlent la présence de zone refuge au cours du Pléistocène dans cette région (Falk *et al*., 2003; Agnèse *et al*., 2003). Pouyaud & Agnèse (1995) travaillant sur la variabilité génétique des populations d'*Oreochromis niloticus* suggèrent que cette espèce est originaire du bassin du Nil à partir duquel elle s'est répandue en Afrique de l'Est et de l'Ouest. Probablement les populations de *C. gariepinus* de cette région ont emprunté la même route. L'un des éléments majeurs qui ont marqués l'évolution des poissons des bassins de l'Est de l'Afrique est la formation du rift au Miocène (Beadle, 1981). Dans cette région on retrouve aussi un modèle fragmenté lié aux évènements du passé. En effet, les populations de *C. gariepinus* sont réparties en deux grands groupes, ceux du Rift Oriental et ceux du Rift Occidentale (Giddelo *et al*., 2002). Les populations du lac Tanganyika et Victoria se retrouvent dans le même groupe du Rift Occidental. Dans cette étude, nous n'avions pas su disposer des échantillons du rift oriental. Néanmoins, les travaux de Giddelo *et al.* (2002) confirment cela. Dans ces régions beaucoup d'auteurs ont pu observer des évolutions locales, liées à une fragmentation du milieu. C'est le cas de l'évolution allopatrique des Cichlidae au lac Tanganyika, Victoria et Malawi (Verheyen *et al*., 2003; Salzburger *et al*., 2005; Koblmüller *et al*., 2008). A part les conséquences du rifting, l'évènement qui a caractérisé cette région est

l'assèchement des lacs au Pléistocène avec comme conséquence l'isolation des poissons dans des refuges. Probablement, le lac Tanganyika a servi de refuge à plusieurs espèces (Lévêque, 1997) dont le *C. gariepinus.* Le genre *Bathyclarias* en est témoin (Agnèse & Teugels, 2005). En Afrique Australe, bassin du Zambèze, les poissons présentent aussi un modèle d'évolution plus fragmenté entre les différents bassins suite au changement des connections entre les différents bassins intervenu à la fin du Miocène et au Plio-Pléistocène (Skelton, 1994). Dans cette région les études de l'évolution des Cichlidae ont révélé la présence des groupes isolés par des barrières (Katongo *et al*., 2005, 2007). Joyce *et al*. (2005) signalent la présence des refuges dans la région lors de l'asséchement du lac Makgadikgadi.

Dans le passé, l'espèce *C. gariepinus* avait été subdivisée en trois espèces avec *C. gariepinus,* Burchell en Afrique de l'Est, *C. mossambicus* Peters en Afrique australe et *C. lazera* Valenciennes en Afrique centrale et de l'Ouest. Les deux dernières espèces sont tombées en synonymies (Ozouf-Costaz *et al.*, 1990). Les résultats obtenus dans cette étude sur la phylogéographie de ce poisson dans le bassin du Congo relance l'opportunité de la séparation une fois de plus en plusieurs espèces suite aux valeurs de support obtenu sur l'arbre phylogénétique, le nombre des mutations et le temps de divergence entre les clades et groupes. Il faudra que la taxonomie tienne compte non seulement de la morphologie mais aussi de la divergence génétique entre les populations, comme c'était le cas du *C. gariepinus* et du *C. anguillaris* (Rognon *et al.,* 1998) et du *C. gariepinus* et *Bathyclarias* sp. (Agnèse & Teugels, 2005).

5.1.2. Etudes phylogéographiques dans le bassin du Congo

Les résultats du chapitre 2 ont révélé la présence de deux clades (Congo et Lulua-Kando) et plusieurs groupes. Le réseau semble mieux indiquer la divergence entre les différents groupes. Dans le chapitre 4, nous avons utilisé plusieurs échantillons en provenance des régions avoisinantes du bassin du Congo. L'arbre phylogénétique ainsi que le réseau clarifient mieux la structure génétique de *C. gariepinus* dans le bassin du Congo. Ainsi, le bassin du Congo possède quatre clades dont trois appartiennent aux bassins qui l'entourent tels le Nil au Nord et Nord-Est, le Zambèze au sud et la région du rift à l'Est. Les études antérieures sur la phylogéographie du poisson chat *C.*

gariepinus n'ont pas été menées à l'intérieur du bassin du Congo sauf dans les zones périphériques telles la région des grands lacs et l'Oubangui (Giddelo *et al.,* 2002, Agnèse & Teugels, 2005). Nos résultats signalent l'occupation de la cuvette centrale pendant le Pléistocène, ce qui corroborent avec la thèse de l'évacuation de l'ancien lac par la mer (Beadle, 1981). L'évolution du chimpanzé bonobo, habitant la foret équatorial, relate la présence du fleuve Congo pendant les périodes arides du Pléistocène (Eriksson *et al.,* 2004), ce qui peut expliquer l'absence de la fragmentation des populations des *Clarias gariepinus* dans cette région comparativement aux populations du bassin Nilo-Soudanique comme proposé par Arndt *et al.* (2003). Les travaux sur les crocodiles dans la même région effectuée par Eaton *et al.* (2009) semble soutenir cette thèse.

La variabilité génétique obtenue dans cette étude (Tableau 3.3.) grâce à l'utilisation des microsatellites semble faible comparativement à celle reportée dans la littérature (cfr Tableau 1.4.). En effet, Galbusera (1997) relate des valeurs variant entre 2,14 à 10,71, 0,13 à 0,78 et 0,18 à 0,79 respectivement pour le nombre d'allèles, l'hétérozygotie observée et attendue. Le faible nombre des loci peuvent être à la base de la situation observée dans notre étude ; beaucoup d'auteurs suggèrent d'utiliser au moins 10 microsatellites (Evanno *et al.,* 2005). Les résultats de Giddelo *et al*., (2002) montrent une diversité des nucléotides variant entre 0.000 et 0.0660. Une deuxième raison pour les faibles diversités obtenues dans le bassin du Congo est probablement liée à la taille de l'échantillonnage.

La divergence entre les populations du *Clarias gariepinus* du Nil et du bassin du Congo soutiennent les hypothèses sur l'état relique. Au sud du Bassin du Congo on retrouve une faune des poissons à affinité élevée avec celle du bassin nord du Zambèze, cas des Cichlidae (*Oreochromis macrochir*, *Serranochromis macrocephalus*) et Clariidae (*Clarias gariepinus, Clarias ngamensis*). Cette étude et les observations de certains auteurs (Bailey, 1986 ; Lévêque, 1997) révèlent qu'il s'agit des poissons reliques ayant occupé la région probablement avant la séparation des différents bassins. Les résultats des études d'Agnèse et Teugels (2005) sur la phylogeographie des poissons chats *C. gariepinus* du Luapula et du lac Malawi, ont pu confirmer cela. Le *C. gariepinus* du bassin du Congo, est aussi une relique, ce qui

confirme les anciennes connexions entre ce bassin et les autres systèmes situés dans le sud, est et nord, et les différentes hypothèses émises dans la littérature pour expliquer la diversité de sa faune (Thieme *et al.*, 2005; Poll, 1963, 1976; Bell Cross, 1966; Lévêque, 1997). D'ailleurs, plusieurs autres espèces du bassin sont repris dans le groupe des reliques, il s'agit par exemple du perche du Nil *Lates niloticus* comme représentant du système Nilo-Sudanique (Lévêque, 1997).

5.1.3. Conservation génétique et durabilité

La faune dans le bassin du Congo est très riche, mais en même temps très fragile. Elle est menacée par de nombreux facteurs anthropiques: croissance de la population, aménagement des infrastructures, extraction de minerais, exploitation forestière, agriculture, pêche, pollution des villes et changements climatiques (Thieme *et al.*, 2005; Katemo, 2009). L'impact de la pollution et la surpêche sur la variabilité génétique des poissons a déjà été signalé dans d'autres régions (Maes *et al*., 2005; Hrbek *et al.,* 2005). Ces deux phénomènes peuvent se traduire par l'augmentation du taux d'homozygotes. L'exploitation forestière modifie l'habitat pour les espèces acclimatées à vivre sous la canopée (Thieme *et al*., 2005). Le *C. gariepinus* n'est pas manifestement une espèce rare et menacée, mais par contre, c'est un bon indicateur de la diversité historique et une riche source de protéine pour la population. Le fait que ce poisson vive dans plusieurs milieux, même dans des eaux moins riches en oxygènes dissous et qu'il possède une grande fécondité, sa disparition d'une rivière peut indiquer un changement anthropogène profond du lieu.

Un autre problème qui guette l'évolution du *Clarias gariepinus* est l'introduction des populations allochtones par l'aquaculture. Ce travail révèle la présence de quatre clades et plusieurs groupes dans le bassin du Congo, certains isolés entre eux, il y a des millions d'années (Upper Lufira (LUF) et Lualaba (DLB)); cela constitue un grand patrimoine génétique (diversité génétique) et historique de l'humanité. Mais cette espèce peut très facilement s'acclimater dans une nouvelle région. En Asie, le *Clarias gariepinus* avait été introduit en aquaculture pour améliorer les performances zootechniques des espèces locales des Clariidae par la technique d'hybridation. Après quelques années, l'évaluation de l'impact de cette introduction sur les populations

sauvages, avait révélé la contamination du pool génétique sauvage par les hybrides (Senanan *et al*., 2004).

5.2. Perspectives

Des études phylogéographiques complémentaires sur d'autres espèces (ex. *Lates niloticus*, *Heterobranchus langifillis*) sont nécessaires pour confirmer cette thèse des reliques. Pour la Cuvette Centrale, l'idéal est de renforcer les études sur la phylogéographie des espèces endémiques à ce milieu. De telles recherches, pourront permettre de comprendre l'évolution de la Cuvette Centrale avant l'écoulement des eaux vers la mer, ainsi que les phénomènes plus récents. En plus sur le haut plateau du Katanga (sud du bassin du Congo) se situe la ligne de faîte qui sépare le bassin du Zambèze à celui du Congo (Poll, 1976). Une étude de la coévolution (*Clarias gariepinus* – parasites) dans cette région peut fournir probablement des éléments intéressants à la compréhension des mécanismes d'évolution et aussi le phénomène de l'ancienne connexion entre les différents bassins. En effet, Barson *et al.* (2010) travaillant sur les parasites du *C. gariepinus* en Afrique de l'Est (Kenya), de l'Ouest (Sénégal) et Australe (Zimbabwe) ont pu observer les phénomènes des anciens contacts entre les différents bassins africains.

References

Adamason, E. (2010). Influence of historical landscapes, drainage evolution and ecological traits on patterns of genetic diversity in Southest Asian freshwater snakehead fishes. Thesis. Qeensland University of Technology. 253 p.

Adeyemo, A. A., A.F. Yakubu, G. A. Oladosu & Ayinla, O.A. (1997). Predation by aquatic insects on African catfish fry. *Aquat. Int.* **5**(1):101-103.

Agnèse, J. F. & Teugels, G. G. (2001a). Genetic evidence for monophyly of the genus *Heterobranchus* and paraphyly of the genus *Clarias* (Siluriformes, Clariidae). *Copeia* **2**: 548-552.

Agnèse, J.-F. & Teugels, G. G. (2001b). The *Bathyclarias- Clarias* species flock. A new model to understand rapid speciation in African Great lakes. *Comptes Rendus de l'Académie des Sciences de Paris*, *Sciences de la vie* **34**: 683-688.

Agnèse, J.-F. & Teugels, G. G. (2005). Insight into the phylogeny of African (Teleostei Siluriformes): implications for their body shape evolution, biogeography and taxonomy. *Molecular Phylogenetics and Evolution* **36**: 546-553.

Allendorf, F.W. & Luikart, G. (2007). Conservation and the genetics of populations. Blackwell Publishing, Malden.

Anene, N. S. & Tianxiang, G. (2007). Is the Dutch domesticated strain of *Clarias gariepinus* (Burchell, 1822) a hybrid? *African Journal of Biotechnology* **6**: 1072-1076.

Anon, P. (1991). Human Foragers in Tropical Rain Forests. *Human-Ecology* **19**(2): 115-285

Anthony, N. M., Johnson-Bawe, M., Jeffery, K., Clifford, S. L., Abernethy, K. A., Tutin, C. E., Lahm, S. A., White, L. J. T., Utley, J. F., Wickings, E. J. & Bruford, M. (2007). The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proceedings of the National Academy of Sciences of the USA* **104**: 20432-20436.

Arndt, A., Van Neer, W., Hellemans, B., Robben, J., Volckaert, F. & Waelkens, M. (2003). Roman trade relationship at Sagalassos (Turkey) elucidated by ancient DNA of fish remains. *Journal of Archeological Sciences* **30**: 1095-1105.

Avise, J. C. (2000). Phylogeography. The history and formation of species (Second edition). *Harvard University Press,* Cambridge, Massachusetts. 447 p

Bailey, R.G. (1986). The Zaïre River system. In *The ecology of rivers systems* (Davies, B. R. & Walker, K. F., eds.), pp. 201-214. Dordrecht: Dr. W. Junk Publishers.

Balloux, F. & Goudet, J., (2002). Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology*. **11**: 771-783.

Banister, K.E. (1986). Fish of the Zaïre system. In *The ecology of rivers systems* (Davies, B. R. & Walker, K. F., eds.), pp. 215-224. Dordrecht: Dr. W. Junk Publishers.

Barson, M., Prikrylova, I., Vanhove, M.P.M, Huyse, T. (2010). Parasite hybridization in African *Macrogyrodactylus* spp. (Monogenea, Platyhelminthes) signals historical host distribution. Parasitology **137**: 1585-1595.

Bartley D.M., Rana, K. & Immink, A.J. (2001). The use of inter-specific hybrids in aquaculture and fisheries. Reviews in *fish Biology and Fisheries,* **10**: 325-337.

Beadle, L.C. (1981). *The inland water of tropical Africa.* London: Longman Group Limited. 455 p.

Belkhir, K.,Borsa, P., Chikhi, L., Raufaste, N., Bonhomme, F. (2004). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier. France.

Bell-Cross, G., (1966). The distribution of fishes in Central Africa. *Fish Research Bulletin of Zambia* **4**: 3–15.

Bermingham, E., McCafferty, S. S. & Martin, A. P. (1997). Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. In: *Molecular Systematics of Fishe*s (Kocher T.D. & Stepien, C.A., eds.), pp. 113-128. San Diego: Academic Press.

Bermingham, E. & Martin, A. P. (1998). Comparative mtDNA phylogeography of neotropical freshwater fishes: testing shared history to infer the evolutionary landscape of lower Central America. *Molecular Ecology* **7**: 499-517.

Benech, V., Teugels, G.G. & Gowne, G. (1993). Critère pratique pour distinguer deux poissons-chats Africains C*larias anguillaris* et *Clarias gariepinus* (siluriformes; Clariidae). *Cybium*, **17**(1): 83-85.

Birky C.W., Feurst, P., Maruyama, T. (1989). Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* **121**: 613-627.

Bowen, B. W., Bass, A. L., Rocha, L. A., Grant, W. S. & Robertson, D. R. (2001). Phylogeography of the trumpetfishes (*Aulostomus*): Ring species complex on a global scale. *Evolution* **55**: 1029-1039.

Briolay, J., Galtier, N., Brito, R. M. & Bouvet, Y. (1998). Molecular phylogeny of Cyprinidae inferred from cytochrome *b* DNA sequences. *Molecular Phylogenetics and Evolution* **9**: 100-108.

Broadley, D. G. & Cotterill, F. P. D. (2004). The reptiles of southeast Katanga, an overlooked 'hot spot'. *African Journal of Herpetology* **53**: 35-61.

Brown, W. M. (1985). The mitochondrial genome of animals. p. 95-130 ***in*** R. H. Macintyre, ed. Molecular evolutionary genetics. Plenum, New York.

Brown, J.H. (1988). Species diversity. In Myers, A.A., Giller, P.S. (Eds.), Analytical biogeography: an integrated approach to the study of animal and plant distributions. Chapman and Hall, London. 57-89.

Brown, D. M., Brenneman, R. A., Koepfli, K-P, Pollinger, J. P, Milà, Georgiadis, N.J, Jr, E. E. L, Grether, G. F., Jacobs, D. & Wayne, R. K. (2007). Extensive population genetic structure in the giraffe. *BMC Biology*, **5**: 57

Brumfield R.T., Beerli P., Nickerson, D.A. & Edwards, S.V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution*, **18**: 249-256.

Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. (2009) TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Cavalli-Sforza L.L. & Edwards A.W.F. (1967). Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics* **19**: 233-257.

Chistiakov D. A., Hellemans, B & Volckaert, F.A.M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture* **255**: 1–29.

Cincotta, R. P., Wisnewski, J. & Engelman, R. (2000). Human population in the biodiversity hotspots. *Nature* **404**: 990-992.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**: 1657-1659.

Clifford, S. L., Anthony, N. M., Joshnson, M B., Abernethy, K. A., Tutin, C. E. G., White, L. J. T., Bermejo, M., Goldsmith, M. L., MCFarland, K., Jeffery, K., Bruford, M. W., & Wickings, E. J. (2004). Mitochondrial DNA phylogeography of western lowland gorillas (*Gorilla gorilla gorilla*). *Molecular Ecology* **13:** 1551-1565.

Cohen, A., Bills, R., Cocquyt., C. Z. & Caljon, A. G. (1993). The impact of sediment pollution on biodiversity in the Lake Tanganyika. *Conservation Biology* **7**(3): 667-677.

Comstock, K. E., Georgiadis, N., Pecon-Slattery, J., Roca, A. L, Ostrander, E. A., O'Brien, S. J. & Wasser, S. K. (2002). Patterns of molecular genetic variation among African elephant populations. *Molecular Ecology* **11**: 2489-2498.

Cornuet J.M. & Luikart G., (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001-2014.

Cottenie, K. & De Meester**,** Luc. (2004). Metacommunity Structure: Synergy of Biotic interactions as selective agents and dispersal as fuel**.** *Ecology* **85**: 114–119.

Cotterill, F. P. D. (2005). The Upemba lechwe, *Kobus anselli*: an antelope new to science emphasizes the conservation importance of Katanga, Democratic Republic of Congo. *Journal of Zoology* **265**, 113-132.

Cotterill, F. P. D. (2006). Taxonomic status and conservation importance of the avifauna of Katanga (south-east Congo Basin) and its environs. *Ostrich* **77**: 1-21.

Dadebo, E**.** (2000). Reproductive biology and feeding habits of the catfish *Clarias gariepinus* (Burchell) (Pisces: Clariidae) in Lake Awassa, Ethiopia. *SINET Ethiop. J. Sci.* **23**(2): 231-246.

Daïnou, K., Bizoux, J-P, J-L Doucet, G Mahy, Hardy, O. J. Haurtz , M. (2010). Forest refugia revisited: nSSRs and cpDNA sequences support historical isolation in a wide-spread African tree with high colonization capacity, *Milicia excelsa* (Moraceae)**.** *Molecular Ecology***. 19** (20): 4462–4477.

De Bont A.F., (1950). Reproduction en étang du *Tilapia melanopleura* (Dum) et *macrochir* (Blgr). Compte rendu de la conférence piscicole anglo-belge, 1949 ; *Communication* **8:** 303-312.

De Graff, G. J. F. & Janssen H. (1996). Artificial reproduction and pond rearing of the African catfish *Clarias gariepinus* in Sub-Sahara. FAO, *Fisheries Technical Paper*. **362**: 73 p.

de Kimpe, P., (1964). Contribution à l'étude hydrobiologique du Luapula-Moero. *Annales du Musée Royale d'Afrique centrale. Tervuren,* **128** : 238 p.

deMenocal, P. B. (2004). African climate change and faunal evolution during Pliocene-Pleistocene. *Earth and Planetary Sciences Letters* **220**: 3-24.

De Vivo, M. & Carmignotto, A. P. (2004). Holocene vegetation change and the mammal faunas of South America and Africa. *Journal of Biogeography* **31**: 943-957.

De Vos, L., Seegers, L., Taverne, L., Thys van den Audenaerde, D., (2001). L'ichtyofaune du bassin de la Malagarasi (Système du LacTanganyika): une synthèse de la connaissance actuelle. *Ann. Mus.R. Afr. Centr. Sci. Zool.* **285**: 117–135.

DeWoody, J. A. & Avise, J. C. (2000). Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology* **56** (3): 461–473.

Domingues, V. S., Bucciarelli, G., Almada, V. C. & Bernardi, G. (2005). Historical colonization and demography of Mediterranean damselfish, *Chromis chromis*. *Molecular Ecology* **14**: 4051-4063.

Doupé, R. G. & Lymbery, A. J. (2000). Managing translocations of aquatic species. *Aquaculture Research* **31**: 151-156.

Ducarme Ch. & J.-C. Micha, (2003). Technique de production intensive du poisson chat africain, *Clarias gariepinus*. *Tropicultura* **21**: 189-198.

Dudgeon, D. (2000). The ecology of tropical Asian rivers and sreams in relation to biodiversity conservation. *Annual Review of Ecology and Systematics* **31**: 239-263.

Eaton, M. J., Martin, A., Thorbjarnarson, J. & Amato, G. (2009). Species-level diversification of African dwarf crocodiles (Genus *Osteolaemus*): Ageographic and phylogenetic perspective. *Molecular Phylogenetics and Evolution* **50**: 496-506.

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

Estoup, A., Gharbi, K., SanCristobal, M., Chevalet, C., Haffray, P., & Guyomard, R. (1998). Parentage assignment using microsatellites in turbot (*Scophthalmus maximus*) and rainbow trout (*Oncorhynchus mykiss*) hatchery populations. Canadian Journal of Fisheries and Aquatic Sciences **55** (3): 715-723.

Eriksson, J., Hohmann, G., Boesch, C. & Vigilant, L. (2004). Rivers influence the population genetic structure of bonobos (*Pan paniscus*). *Molecular Ecology* **13**: 3425-3435.

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611-2620.

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3:** 87-112.

Falk, T M., Teugels, G. G., Abban, Eddie K., Villwock, W. & Renwrantz, L. (2003). Phylogeographic patterns in populations of the black-chinned tilapia complex

(Teleostei, Cichlidae) from coastal areas in West Africa: support for the refuge zone theory. Molecular Phylogenetics and Evolution **27**: 81-92.

Fu, Y.X. (1997). Statistical test of neutrality of mutation against population growth, hitchhiking and background selection. *Genetics* **147**, 915-925

FAO (2007). The value of Biodiversity. *Nature and Faune*. **22** (1). 53p.

FAO (2010).http://www.fao.org/fishery/topic/13533/en

Felip, A., Young, W.P., Wheeler, P.A. & Thorgaard, G.H. (2005). An AFLP-based approach for the identification of sex-linked markers in rainbow trout (*Oncorhyncus mykiss*). *Aquaculture*. **247**: 35-43.

Galbusera, P., Volckaert, F.A., Hellemans, B.; Ollevier. F.(1996). Isolation and characterization of microsatellite markers in the African catfish, *Clarias gariepinus* (Burchell, 1822). *Molecular Ecology* **5**: 703–705.

Galbusera, P. (1997). The genetic variability of wild and inbred population of the African catfish *Clarias gariepinus* (Burchell, 1822). Thesis, K.U. Leuven.

Gasse, F. (2000). Hydrological changes in the African tropics sinces the Last Glacial Maximum. *Quaternary Science Reviews* **19**: 189-211.

Giddelo, C. S., Arndt, A. D. & Volckaert, F. A. M. (2002). Impact of rifting and hydrography on the genetic structure of *Clarias gariepinus* in eastern Africa. *Journal of Fish Biology* **60**: 1257-1266.

Goldstein D.B., Linears A.R., Cavalli-sforza, L.L. and Feldman, M.W. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463-471

Greenwood, P. A. (1993). Review of the Serranochromine cichlid fish genera, Pharyngochromis, Sargochromis,Serranochromis and Chetia. *Bull. Nat. Hist. Mus. Zool. Ser.* **59**: 33–44.

Guindon, S, & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696-704.

Guindon, S, Dufayard, J.F, Hordijk, W, Lefort V & Gascuel, O. (2009). PhyML: fast and accurate phylogeny reconstruction by maximum likelihood. *Infection Genetics and Evolution*, **9**: 384-385.

Haffer J. (2008). Hypotheses to explain the origin of species in Amazonia**.** *Braz J Biol.* **68**(4):917-47.

Hardy O.J., charbonnel N., freville H. & Heuertz M. (2003) Microsatellite allele sizes: A simples test to assess their significance on genetic differentiation. *Genetics*, **163**: 1467-1482.

Hardy, O.J. & Vekemans X. (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2:** 618-620.

Hartl, D.L. and Clark, A.G. (1997). Principles of Population Genetics (Third Edition). *Sinuauer Associations, Inc. Publishers Sundeland, Massachusetts.*

Haylor, G. S. (1993). Controlled hatchery production of African catfish *Clarias gariepinus* (Burchell, 1822): an overview. *Aquacult. Fish. Manage* **24**: 245-252.

Hecht, T, Uys, W & Britz, PJ, (1988). The culture of sharptooth Catfish, *Clarias gariepinus* in Southern Africa, South African. National scientific programme report **153**: 133p.

Hecht T., Lublinkhof W. & Kenmuir, D. (1996). Induced spawning of vundu *Heterobranchus longifilis*, and embryo survival rates of pure and reciprocal clariid crosses. *S. Afr. J. Wildl. Res*. **21**: 123-125.

Hewitt, M. G. (2004). The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology* I **4**: 1-16.

Hogan, Z.,. Baird, I. G, Radtke R. & Zanden, M. J. V. (2007). Long distance migration and marine habitation in the tropical Asian catfish, *Pangasius krempfi. Journal of Fish Biology.* **71** (3): 818–832.

Hogendoorn, H. (1983). Growth and production of the African Catfish *Clarias gariepinus* (C and V): III. Bioenergetics relations of body weight and feeding level. *Aquaculture* **35**: 1-17.

Hrbek, T., Farias, I.P., Crossa, M., Sampaio, I., Porto, J.I.R., Meyer, A., (2005). Population genetic analysis of *Arapaima gigas*, one of the largestfreshwater Wshes of the Amazon basin: implications for its conservation. *Anim. Conserv*. **8**, 297–308.

Hubert, N. & Renno, J-F. (2006). Historical biogeography of South American freshwater fishes. *Journal of Biogeography* **33**, 1414–1436

Hubert, N., Duponchelle, F., Nuñez, J., Garcia-Davila, C., Paugy, D. & Renno, J.-F. (2007). Phylogeography of the piranha genera *Serrasalmus* and *Pygocentrus*: implications for the diversification of the Neotropical ichthyofauna. *Molecular Ecology* **16**: 2115-2136.

Huelsenbeck, J. P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **7**: 754-755.

Hughes, R. H. & Hughes, J. S. (1992). *A Directory of African Wetlands.* Gland, Nairobi, and Cambridge: IUCN, UNEP and WCMC.

Huisman, E. A. & Richter, C. J. J. (1987). Reproduction, growth, heath control and aquacultural potential of the African catfish, *Clarias gariepinus* (Burchell, 1822). *Aquaculture*. **63**, 1-14.

Islam, M. N. Islam, Md. S & Alam, Md. S. (2007). Genetic Structure of Different Populations of Walking Catfish (*Clarias batrachus* L.) in Bangladesh. *Biochemical Genetics* **45** (9-10) : 647-662.

Janssen, J.A.L. (1985). Elevage du poisson-chat Africain *Clarias lazera* (int. Et al. 1840) en République Centre Africaine, Alevinage en écloserie, FAO, Rome, *Document technique* **21** : 52 p

Jansen, G., Devaere, S. Weekers, P.H.H. & Adriaens, D. (2006). Phylogenetic relationships and divergence time estimate of African anguilliform catWsh (Siluriformes: Clariidae) inferred from ribosomal gene and spacer sequences Molecular. *Phylogenetics and Evolution* **38** : 65–78.

Joanne A. Labate, Larry D. Robertson, Feinan Wu, Steven D. Tanksley, and Angela M. Baldo. (2009). EST, COSII, and arbitrary gene markers give similar estimates of nucleotide diversity in cultivated tomato (Solanum lycopersicum L.). *Theoretical and Applied Genetics* **118** (5): 1005-1014.

Johnson T. C., Scholz C. A., Talbot M. R., Kelts, K., Ricketts, R. D., Ngobi, G., Beuning, K., Ssemmanda, I. & McGill, J. W. (1996). Late Pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* **273**: 1091-1093.

Johnson, M. B., Clifford, S. L., Goossens, B., Nyakaana, S., Curran, B., White, L. J. T., Wickings, E. J. & Bruford, M. W. (2007). Complex phylogeographic history of central African forest elephant and its implications for taxonomy. *BMC Evolutionary Biology* **7**: 244.

Joyce, D. A., Lunt, D.H., Bills, R., Turner, G. F. ; Katongo, C., Duftner, N., Sturmbauer, C. & Seehausen, O. (2005). An extant cichlid fish radiation emerged in an extinct Pleistocene lake. *Nature* **435**: 90–95.

Kamdem Toham, A., D'Amico, J., Olson, D., Blom, A., Trowbridge, L., Burges, N., Thieme, M., Abell, R., Caroll, R.W., Garlan, S., Langrand, O., Mikala Mussavu, R., O'Hara, D. & Stand, H. (2003). Biological priorities for conservation in the Guinean-Congolian forest and freshwater region. WWF-CARPO, Libreville, Gabon.

Katemo, M., B. (2009). Evaluation de la contamination de la chaine trophique par les élements tracesmétalliques dans le bassin de la Lufira supérieur (Katanga/RD Congo). Mémoire de troisième cycle en Biologie et environement. Unilu

Katongo, C., Koblmüller, S., Duftner, N., Makasa, L. & Sturmbauer, C. (2005). Phylogeography and speciation in the *Pseudocrenilabrus philander* species complex in Zambian Rivers. *Hydrobiologia* **42**: 221-233.

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

Koblmüller, S., Sturmbauer, C., Verheyen, E., Meyer, A. & Salzburger, W. (2006). Mitochondrial phylogeny and phylogeography of East African squeaker catfishes (Siluriformes: *Synodontis*). *BMC Evolutionary Biology* **6**: 49.

Koblmüller, S., Schliewen, U. K., Dufner, N., Sefc, K. M., Katongo, C. & Sturmbauer, C. (2008). Age and spread of the haplochromine cichlid fishes in Africa. *Molecular Phylogenetics and Evolution* **49**: 153-169.

Küper W., Sommer J.H, Lovett J.C., Mutke, J., Linder, H.P., Beentje, H.J., Van Rompaey, R.S.A.R., Chatelain, C., Sosef, M. & Barthlott, W. (2004). Africa's hotspots of biodiversity redefined. *Annals of the Missouri Botanical Garden* **91**: 525–536.

Lal, K.K.; Singh, R.K.; Mohindra, V. Singh, B. et Ponniah, A.G., (2003). Genetic Make Up of Exotic Catfish *Clarias gariepinus* in India. *Asian Fisheries Science* **16**: 229-234.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Larmuseau, M. H. D., Van Houdt, J. K. J., Guelinckx, J., Hellemans, B. & Volckaert, F. A. M. (2009). Distribution and demographic consequences of Pleistocene climate fluctuations for a marine demersal fish in the North-Eastern Atlantic. *Journal of Biogeography* **36**: 1138-1151.

Larsson, L.L., Laikre, L., Palm, S., Andre., C., Carvalho, G. R. and Ryman, N. (2007). Concordance of allozyme and microsatellite differentiation in a marine fish, but evidence of selection at a microsatellite locus. *Molecular Ecology* **16**: 1135–1147.

Laturbene, E. M., Slevaux, J.C. & Sinha, R. (2005). Tropical Rivers. *Geomorphology* **70**: 185-206.

Lévêque, C. (1997). Biodiversity dynamics and conservation: the freshwater fish of tropical Africa. Cambridge: *Cambridge University Press*. 438 p.

Lévêque, C. & Paugy, D. (2006). Distribution géographique et affinités des poissons d'eau douce africains, pp. 59-74. *In* : Lévêque C. & Paugy D., éds. *Les poissons des eaux continentales africaines: diversité, écologie et utilisation par l'homme*. IRD, Paris, 564p.

Librado, P. & Rozas. J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451-1452.

Liu, Z.J. & Cordes J.F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**: 1 –37

Lorenzen, E. D., Arctander, P & Siegimsund. (2006). Regional Genetic Structuring and Evolutionary History od the Impala *Aepyceros mempus*. *Journal od Heredity*. **97** (2): 119-132.

Maes, G.E., Raeymaekers, J.A.M., Pampoulie, C., Seynaeve, A.Goemans, G., Belpaire, C., Volckaert, F.A.M. (2005). The catadromous European eel *Anguilla anguilla* (L.) as a model for freshwater evolutionary ecotoxicology: Relationship between heavy metal bioaccumulation, condition and genetic variability. *Aquatic Toxicology*, **73**: 99–114.

Magis, N. (1961). Nouvelle contribution à l'étude hydrolobiologique des lacs de Mwadingusha, Koni, Nzilo. Liège: F.U.L.R.E.A.C.

Maley, J. (2001). The impact of arid phases on the African rain forest through geological history. In *African Rain Forest Ecology and Conservation. An interdisciplinary perspective* (Weber, W., White, L., Vedder, A. & Naughton-Treves, L., eds.), 68-87. New Haven: Yale University Press.

Mantel N. (1967). The detection of desease clustering and generalized regression approach. *Cancer Research* **27**: 209-220.

Maslin, M. A. & Christensen, B. (2007). Tectonics, orbital forcing, global climate change, and human evolution in Africa: introduction to the African paleoclimate special volume. *Journal of Human Evolution* **53**: 443- 464.

Matthes H. (1964). – Les poissons du lac Tumba et de la région d'Ikela. *Annales du Musée. Royale de Afrique centrale. Sciences. Zoologiques,* **126:** 204 p.

Mbega, J.D., et Teugels, G.G., (2003); Guide de détermination des poissons du Bassin inférieur de l'Ogoue, *Presse universitaire de NAMUR*, 165 p.

Mhlanga, W. (2003). Food and feeding habits of tigerfish, *Hydrocynus vittanus*, in Lake Kariba, Zimbabwe. p. 76-81. In: M.L.D. Palomares, B. Samb, T. Diouf, J.M. Vakily and D. Pauly (eds.) Fish biodiversity: local studies as basis for global inferences. *ACP-EU Fish. Res. Rep.* **14**: 281p.

Micha J.-C. (1973). Etude des populations piscicoles de l'Oubangui et tentative de sélection et d'adaptation de quelques espèces à l'étang de pisciculture. Centre Technique Forestière tropicale, Norgent sur Marne, 100 p.

Moodley, Y. & Bruford, M. W. (2007). Molecular Biogeography: Towards an Integrated Framework for Conserving Pan-African Biodiversity. *PloS ONE* **2**,

e454

Mohindra, V., Singh,, R. K. M. Palanichamy, Ponniah, A. G. and Lal, K. K. (2007). Genetic identification of three species of the genus Clarias using allozyme and mitochondrial DNA markers. *J. Appl. Ichthyol.* **23**: 104–109.

Morin, P. A., Martien, K. K., and Taylor, B. L. (2009) Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* **9** (1): 66-73.

Morin, P.A., Luikart, G. & Wayne, R.K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, **19**: 208-216.

Murienne, J. (2009). Testing biodiversity hypotheses in New Caledonia using Phylogenetics. *Journal of biogeography* 36, 1433-1434.

Mwita, C. J. & Nkwengulila, G. (2008). Molecular phylogeny of the clariid fishes of Lake Victoria, Tanzania, inferred from cytochrome *b* DNA sequences. *Journal of Fish Biology* **73**: 1139-1148.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**: 853-858

Nan-Nakorn, U., Taniguchi, N., Nugroho, E., Seki, S & Kamonat, W. (1999). Isolation and characterization of Microsatellite Loci of Clarias macrocephalus to Genetic Diversity Study. Fisheries Sciences. **65** (4): 520-526.

Nan-Nakorn, U., Sodsku, P., Wongrat, P., Janekitkarn, S & Barrley, D.M. (2002). Izozyme variation among four species of the catfish genus *Clarias*. *Journal of Fish Biology*. **60:** 1051–1057.

Narum, S. R., Banks, M. , Beacham, T. D., Bellinger, M. R., Campbell, M. R., Dekoning, J., Elz, A., Guthrie, C. M., Kozfkay, C., Miller, K. M., Moran, P., Phillips, R., Seeb, L. W., Smith, C. T., Warheit, K., Young, S. F. and Garza, J. C. (2008). Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology* **34**: 3464- 3477.

Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University

Press.

Nei, M. & Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**: 145-163.

Nersting, L. G. & Arctander, P. (2001). Phylogeography and conservation of impala and greater kudu. *Molecular Ecology* **10**: 711-719.

Nwadukwe, F.O., (1995). Early growth and survival of clarias gariepinus (Burchell), Heterobranchus longifilis (Val) (Pisces: Calriidae) and their F_1 hybrids in ponds. Netherlands *Journal of Aquatic Ecology* **29** (2): 177-182.

O' Connell, M. & Wright, J.M. (1997). Microsatellite DNA in fishes. *Reviews of Fish Biology and Fisheries* **7**: 331-363.

Ozouf-Costaz, C., Teugels, G. G. & Legendre, M. (1990). Karyological analysis of three strains of the African catfish, *Clarias gariepinus* (Clariidae), used in aquaculture. *Aquaculture* **87**: 271-277.

Palti, Y. (2009). Aquaculture Genomics. In Molecular research in aquaculture, Overturf, K, ed, *Library of Congress Cataloging-in-Publication Data.* 395p.

Pasyanos, M. E. & Nyblade, A. A. (2007). A top to bottom lithospheric study of Africa and Arabia. *Tectonophysics* **444**: 27–44.

Pemberton, J.M., Slate J. & Bancroft, D.R., Barrett, J.A. (1995). Non-amplified alleles at microsatellite loci: a caution for parentage and populations studies. *Molecular Ecology* **4**: 249-252.

Plana, V. (2004). Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest. *Philosophical Transactions of the Royal Society of London B* **359**: 1585-1594.

Poll, M., (1959). Aspects nouveaux de la faune ichtyologique du Congo Belge, *Bull. Sc. Zool. Fr.*, Tome LXXXIV **4** : 259 – 271.

Poll, M. (1963). Zoogéographie ichtyologique causale du Katanga. *Université de l'Etat. Publications C* : 95-106.

Poll, M. & Gosse, J.P., (1963). Contribution à l'étude systématique de la faune ichtyologique du Congo central, *Ann. Mus. Roy. Afr. Cent.* **116** : 45 – 110.

Poll, M. (1976). Poissons: exploitation du Parc National de l'Upemba. *Fondation pour Favoriser les Recherches Scientifiques en Afrique* **73**: 1- 127.

Poll, M., 1986. Classification des cichlidae du lac Tanganyika: tribus, génères et espèces. *Mem. Aca. R. Belgique* (Science) **45:**1-163.

Poompuang, S. & Na-Nakorn, U. (2004). A preliminary genetic map of walking catfish (Clarias macrocephalus). *Aquaculture* **232**: 195–203.

Posada, D. (2008). jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution* **25**: 1253-1256.

Posada, D. (2009). Selection of models of DNA evolution with jModelTest. In *Bioinformatics for DNA sequence analysis* (Posada, D., ed.). *Methods in Molecular Biology* **537**: 93 -112.

Pouyaud, L. & Agnèse, J.-F. (1995). Phylogenetic relationships between 21 species of three tilapiine genera *Tilapia, Sarotherodon* and *Oreochromis* using allozyme data. *Journal of Fish Biology.* **47 (1)**: 26-38.

Pritchard J.K., Stephens M & Donneley, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**, 954-959.

Rainboth, W. J. (1996). Fishes of the Cambodian Mekong. FAO, ROME, 263 p

Ramos-Onsins S. E. & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**: 2092-2100.

Rand, D. M. 1996 Neutrality tests of molecular markers and the connection between DNA polymorphism, demography and conservation biology. *Conserv. Biol*. 10, 665-671.

Renaut, S., Nolte, A W. and Bernatchez, L. (2010). Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus spp.* Salmonidae). *Molecular Ecology* **19** (1): 115–131.

Renno, J.-F. , Hubert, N. , Torrico, J.P., Duponchelle, F. , Rodriguez, J. N., Davila, C. G., Willis, S.C. & Desmarais, E. (2006). Phylogeography of *Cichla* (Cichlidae) in the upper Madera basin (Bolivian Amazon). Molecular Phylogenetics and Evolution **41**: 503–510.

Roark, S. & Brown, K. (2009). Effects of metal contamination from mine tailings on allozyme distributions of populations of great plains fishes. *Environmental Toxicology and Chemistry*. **15** (6): 921–927.

Roberts, T. R., (1975). Geographical distribution of African freshwater fishes. *Zoological Journal of the Linnean Society* **57**: 239-317.

Roberts, T.R. & Stewart, D.J., (1976). An Ecological and Systematic survey of Fishes in rapids of the lower Zaïre or Congo River, *Bull. Mus. Comp. Zool.,* **147** (6): 239-317.

Rognon, X., Teugels, G. G, Guyomard, R., Galbusera, P., Andriamanga, M., Volckaert, F. A. M. & Agnèse, J. F. (1998). Morphometric and allozymes variation in the African catfishes *Clarias gariepinus* and *C. anguillaris*. *Journal of Fish Biology* **53**: 192-207.

Rognon, X & Guyomard, R. (2003). Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa. *Molecular Ecology* **12** (2): 435–445.

Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3. Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574.

Ruzzante, D. E., Mariani, S., , Bekkevold, D., , André, C., Mosegaard, H., Clausen, L. A.W, Dahlgren, T. G., Hutchinson, W. F., Hatfield, E. M.C, Torstensen, E., Brigham, J., Simmonds,J.E., Laikre, L., Larsson, L. C., Stet, R. J.M, Ryman, N. and Carvalho, G. R. (2006). Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. *Proceedings the Royal Society B*. 273: 1459-1464.

Sahoo, S.K. ; Giri, S.S. ; Sahu, A.K. & Ayyappan, S., (2003). Experimental Hybridization between Catfish *Clarias batrachus* (Linn.) x *Clarias gariepinus* (Bur.) and Performance of the Offspring in Rearing Operations. *Asian Fisheries Science* **16**: 157-166.

Salzburger, W., Mack, T., Verheyen, E. & Meyer, A. (2005). Out of Tanganyika: Genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology* **5:** 1-15.

Sauvage, C., Bierne , N. Lapègue, S., Boudry, P. (2007). Single Nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas.* Gene **406**, 13–22.

Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002). TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502-504.

Ségalen, L,. Lee -Thorp, J. A. & Cerling, T. (2007). Timing of C_4 grass expansion across sub-Saharan Africa. *Journal of Human Evolution* **53**: 549-559**.**

Senanan,W., Kapuscinski, A.R., Na-Nakorn, U, & Miller, L.M. (2004). Genetic impacts of hybrid catfish farming (*Clarias macrocephalus x C. gariepinus*) on native catfish population in central Thailand. *Aquaculture*. **235:**167-184.

Shawn R. Narum, Jon E. Hess, and Andrew P. Matala. (2010) Examining Genetic Lineages of Chinook Salmon in the Columbia River Basin. *Transactions of the American Fisheries Society* **139** (5): 1465-1477.

Skaltin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**: 457- 462

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

Skelton, P. H. & Teugels, G.G. (1992). Neotype description for the African Catfish *Clarias gariepinus* (Burchell, 1822) (Pisces: Siluroidei: Clariidae). *Ichtyological Bulletin of the J.L.B. Smith Institute of Ichthyology* **56**: 1-8.

Smith, C. T., Antonovich, A., Templin, W. D., and Elfstrom, C. M., Narum S. R. Seeb, Lisa, W. (2007). Impacts of Marker Class Bias Relative to Locus-Specific Variability on Population Inferences in Chinook Salmon: A Comparison of Single-Nucleotide Polymorphisms with Short Tandem Repeats and Allozymes. *Transactions of the American Fisheries Society* **136**, 1674-1687.

Snoeks, J., De Vos, L. & Thys van den Audernaerde, D. (1997). The ichthyo-geography of Lake Kivu. *South African Journal of Science* **93**: 1-6.

So, N, Van Houdt, J. K J & Volckaert, F. A .M. (2006). Genetic diversity and population history of the migratory catfishes *Pangasianodon hypophthalmus* and

Pangasius bocourti in the Cambodian Mekong River. *Fisheries Sciences* **72**: 469-476.

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.

StatSoft. Inc. (2009). STATISTICA (data analysis software system), version 9.0. w.w.w. Statsoft.com

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

Sturmbauer, C., Koblmüller, S., Sefc, K.M. & Duftner, N. (2005). Phylogeographic history of the genus Tropheus, a lineage of rock-dwelling cichlid fishes endemic to Lake Tanganyika. *Hydrobiologia* **542**: 335–366.

Swofford, D.L. (2002). PAUP*. Phylogenetic analysis using parsimony (*and other methods), Version 4.0b10. Sunderland: Sinauer Associates.

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

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Molecular Biology and Evolution* **9**: 678-687.

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

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. In *Some mathematical questions in biology – DNA sequence analysis* (Miura, R. M., ed.), pp. 57-86. Providence: American Mathematical Society.

Teugels, G. G. (1986). A systematic revision of the African species of the genus *Clarias* (Pisces; Clariidae). *Annales du Musée Royal de l'Afrique Centrale* **247** : 199 p.

Teugels, G. G, (2003). State of the art of recent Siluriform systematics. In: Arratia, G., Kapoor, B.G., Chardon, M., Diogo, R. (Eds.), CatWshes, vol. 1. Science Publishers, Inc., EnWeld, USA, 317–352.

Thieme, M. L, Abell, R, Stiassny, M. L. J, Skelton, P., Burgess, N. D., Lehner, B., Dinerstein, E., Olson, D., Teugels, G. G. & Kamdem-Toham, A. (2005). *Freshwater Ecoregions of Africa and Madagascar. A Conservation Assessment.* Washington DC: Island Press. 483p.

Thompson, J. D.; Gibson, T. J.; Plewniak, F.; Jeanmougin, F. & Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876-4882.

Thys van den Audenaerde, D. F. E. (1964). Révision systématique des espèces congolaises du genre *Tilapia* (Pisces, Cichlidae). *Annales du Musée Royal de l'Afrique Centrale, Sciences Zoologiques.* 124 p.

Tosi, A. J. (2008). Forest monkeys and Pleistocene refugia: a phylogeographic windows on to disjunct distribution of the *Chlorocebus thoesti* species group. *Zoological Journal of the Linnean Society* **154**: 408- 418.

Trauth, M. H, Larrasoana, J. C. & Mudelese, M. (2009). Trends, rhytms and events in Plio-Pleistocene African climate. *Quaternary Sciences Reviews* **28**: 399-411.

Quérouil, S., Verheyen, E., Dillen M. & Colyn, M. (2003). Patterns of diversifications in two African forest shrews: *Sylvisorex johnstoni* and *Sylvisorex ollula* (Soricidae, Insectivora) in relation to paleo-environmental changes. *Molecular Phylogenetics and Evolution* **28**, 24-37.

Van der Bank, F. H., Grobler, J. P., & du Preez, H. H. (1992). A comparative biochemical genetic study of three populations of domesticated and wild African catfish (*Clarias gariepinus*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **101**: 387-390.

Van der Walt, L.D., Van der Bank, F.H. & Steyn, G.J. (1992). Allozyme variation in domesticated Africancatfish *Clarias gariepinus*. from The Netherlands. *Comparative Biochemistry and Physiology* **104**: 15–18.

Van der Walt, L.D., Steyn, G.J. & Van der Bank, F.H., (1993). The suitability of using cryopreservation ofspermatozoa for the conservation of genetic diversity in African catfish *Clarias gariepinus*.. *Comparative Biochemistry and Physiology* **106**: 313–318.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. (2004). Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**: 535–538.

Verheyen, E., Salzburger, W., Snoeks, J. & Meyer, A. (2003). Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science* **300:** 325-329.

Verreth, J.A.J. & Eding E.H. (1993). European farming industry of African catfish (*Clarias gariepinus*): facts and figures. *World Aquaculture'93, Torremolinos, Spain*, 6-13.

Viveen, W.J.A.R.; Richter, C. J. J., Van Dordet, P. G. W. J., Janssen, J. A. L. & Huisman, E. A., (1985). Manuel pratique de pisciculture du poisson-chat africain (*Clarias gariepinus*). La Haye.

Volckaert, F.A.M. & Hellemans, B. (1999). Survival, growth and selection in a communally reared multifactorial cross of African catfish *Clarias gariepinus*. *Aquaculture*, **171**: 49–64.

Vanstaen, Dieter & Ollevier, Frans. (1994). Gynogenesis in the African catfish (*Clarias gariepinus*). I. Induction of meiogynogenesis with thermal and pressure shocks. *Aquaculture* **128** (3-4): 221-233.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R & Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature* **467:** 555-561.

Wachirachaikarn, A., Rungsin, W., Srisapoome, P. & Na-Nakorn, U. (2009). Crossing of African catfish, *Clarias gariepinus* (Burchell, 1822), strains based on strain selection using genetic diversity data. *Aquaculture* **290:** 53-60.

Wamuini Lunkayilakio, S., Vreven, E., Vandewalle, P., Mutambue, S. & Snoeks, J. (2010). Contribution à la, connaissance de l'ichtyofaune de l'Inkisi au Bas-Congo (R.D. du Congo). *Cybium* **34**: 83-91.

Weir B.S. & Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 68, 1358-1370.

Więcaszek, B., Krzykawski, S., Antoszek, A., Kosik, J. & Serwotka, P. (2010). Morphometric characteristics of the juveniles North African catfish Clarias

gariepinus (Burchell, 1822) from the Heated water aquaculture. *Electronic Journal of Polish Agricultural Universities* **13** (2).

Wilson, A. C., Cann, R. L, Carr, S. M., George, M., Gyllensten, U. B., Helmbychowski, K. M., Higuchi, R. G., Palumbi, S. R., Prager, E. M., Sage, R. D. & Stoneking, M. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society* **26** (4): 375–400.

Wilson, A. B., Teugels, G.G. & Meyer, A. (2008). Marine Incursion: The Freshwater Herring of Tanganyika Are the Product of a Marine Invasion into West Africa. PLoS ONE **3**(4): e1979. doi:10.1371/journal.pone.0001979.

Yalçin, S., Akyurt, I. & Solak, K. (2001). Stomach Contents of the Catfish (*Clarias gariepinus* Burchell, 1822) in the River Asi (Turkey) *Turk. J. Zool* **25**: 461-468.

Yue, G.H., Kovacs, B. & Orban, L. (2003). Microsatellites from *Clarias batrachus* and their polymorphism in seven additional catfish species. *Molecular Ecology Notes* **3**:465–468.

Zamba, A. I. & Vreven, E. (2008). *Rhabdalestes yokai* (Characiformes: Alestidae), a new species from the Léfini and Sangha River basins (Congo River basin), Africa. *Ichthyol. Explor. Freshwaters*, **19** (4): 377-384.

Zhang, D. X. & Hewitt, G. M. (1996). Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution* **11**: 247-251.