Population structure and gene flow in desert environments: an application of molecular tools to isolated fish populations in West Africa

Jolita Dilytė
Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia
2014

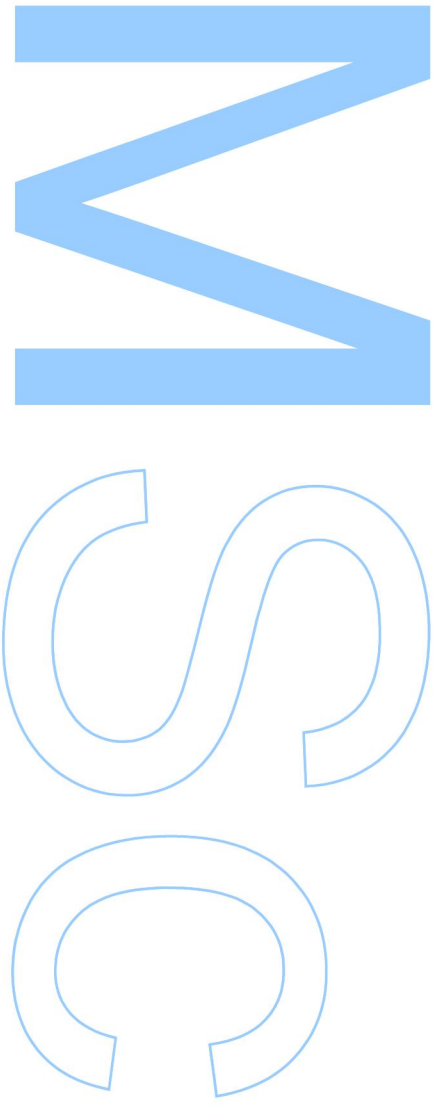
Orientador
Paulo Jorge de Barros Alexandrino, Associate Professor, CIBIO-InBIO/FC-UP

Coorientador
José Carlos Alcobia Rogado de Brito, Senior Scientist, Associate Researcher, CIBIO-InBIO/FC-UP

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



“Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.”

Marie Curie

Acknowledgements

First of all, I would like to express my warm thank to my supervisor Professor Paulo Alexandrino for the opportunity to work in his team and to grow as biologist scientist. All of this has influenced my love to fishes, as to the most interesting and challenging organism to be studied.

I also want to thank to my co-supervisor Doctor José Carlos Brito for his expert advices, monitoring, collection of samples in the trips to the harsh Africa environments, his knowledge about study area, for the help fighting with my fears and the opportunity to be part of his team.

I am very thankful to Doctor Raquel Godinho for her expert advices, sharing the knowledge and encouragement and supporting words throughout difficulties during this thesis.

I would like to thank Doctor Stephen Sabatino for his guidelines, monitoring, help and encouragement throughout this thesis.

My appreciation to Susana Lopes for her help in the lab and analyzing the microsatellites, sequences, for giving advises and support.

I am also very thankful to the members of the Biodeserts team for their friendship and all help through all study process: trying to paint the view of the study area in my mind, dealing with different programs and software, giving their advices and collection of the samples in the field trips.

Also for the samples collection I would like to thank to Camilo Carneiro, Edna Correia, Miguel Clavero.

I also express my sincere gratitude to many people: my colleges and Professors of the Master thesis course, PhD students, Post-doc and technicians of CTM and CIBIO, for all their advices in the lab, using different kind of programs and software, sharing their knowledge, support, for having brainstorming during and after work.

All these people during this process made me better scientist and even better person. Only good begets good and only sharing information and working as a team we can make our planet better place to live! Thank you for one of the best events in my life!

The last but not the less, a huge thanks to my family, boyfriend and friends for all their support, love and forgiveness for missing important events, and just believing in me and being always for me. In the end I would like to special

appreciation to my boyfriend to all those sleepless nights and my bad mood and for always believing in me.

Thank you ALL!!

Abstract

The Sahara desert fluctuated between desert and tropical conditions over the past seven million years, which has had a profound effect on the evolution of its flora and fauna. Desertification of the Sahara appears to have promoted species diversification in some cases. This diversification is probably due to the repeated isolation of lineages in extreme environmental conditions, which can promote ecological adaptation and divergence.

The diversification of species via ecological adaptation seems to be particularly important in aquatic taxa that are distributed in lowland and mountain areas, as they are often isolated and connected through changes in stream and river drainages. Changes in precipitation in desert areas are often extreme and occur seasonally as well as decadal and much longer scales (e.g., the Pleistocene glaciations). Consequently, many relict fish populations survived in those habitats with extreme climates.

The main goal of this research was to measure population structure and genetic diversity in Tilapia fishes from the mountain and lowland regions of Mauritania. We chose to study Tilapias to help to determine if the desertification of the Sahara has impacted rates of divergence of aquatic species by altering patterns of dispersal and variance and promoting ecological adaptation. The Tilapia specimens collected for this research were obtained from rivers and streams across North-West Africa, including both coastal and inlands areas with a focus on locations within Mauritania. Molecular phylogenetic analysis was used to help determine the species to which each of our samples was a member. Specifically, were conducted a phylogenetic analyses of collected specimens using mitochondrial DNA (16S and ND2) and nuclear DNA (nDNA) (1st intron of S7) sequence data. The results showed that Tilapia samples consisted of individuals from two tribes: Oreochromini and Coptodonini. Each tribe contained two genera: *Sarotherodon (*A and B lineages*)* and *Coptodon (*A and B lineages*)*, with their distribution in coastal (A) and inland (B) waters. Mauritania inland waters contained the members of *Sarotherodon* (B) and *Coptodon* (B) genera, while *Sarotherodon* B was predominant in mountains and lowlands. The population-level analyses were focused on the individuals of *Sarotherodon* A and B lineages, due to their close genetic relationship based on the nuclear locus used for phylogenetic analyses.

Thirteen novel microsatellite loci were developed to answer the main study question. Six statistically significant genetic groups (F_{ST} range: from 0.1209 to 0.6413, P<0.005 after Bonferroni correction) were found, two of which were genetically differentiated populations: Morocco and coastal Mauritania. While the rest four groups were identified within Mauritania inland. Two groups out of four had broad and sympatric distribution, and were distinct from the rest of Mauritanian populations at different genetic levels.

The main study finding was the Affolé mountain population isolation from the western Mauritania populations (F_{ST} 0.1209). The Affolé population did not exhibit any private alleles, which could be a signal of the recent bottleneck. Tilapia population dynamics appear to generally be related to the hydraulic network in each sub-basin but still sample size disallowed us to formerly test this hypotheses. Besides, in Karakoro sub-basin were found individuals from different populations. The results of this study help to understand the phylogenetic relationships among Tilapias in North and West Africa. The diversification of Tilapias appears to be related to their isolation in extreme environments. The isolation of lineages of Tilapia may be due to their dispersal abilities and patterns of connectivity in regions where they are found, which are sensitive to changes in precipitation on yearly and longer time scales.

Key word: African cichlids; phylogenetics; population genetics; genetic structure; hydraulic networks.

Table of Contents

List of figures

List of tables

Abbreviations

A_R - Allelic richness
BI - Bayesian inference
bp - Base pair
°C - Celsius degrees
ddH2O - Double-distilled water
DNA - Deoxyribonucleic acid
dsDNA – Double-stranded DNA
F_{IS} - Inbreeding coefficient
F_{ST} – Fixation index of genetic differentiation
H - Haplotype
H_E - Expected heterozygosity
H_O - Observed heterozygosity
HWE - Hardy-Weinberg equilibrium
K - Number of population assumed by STRUCTURE
km - Kilometer
Met - Methionine
min - Minute
MJ - Median-joining algorithm
ML - Maximum likelihood
mtDNA - Mitochondrial DNA
Mya - Million years ago
n - number of individuals/ other units
Na - Number of alleles per locus
ND2 - Mitochondrially encoded NADH dehydrogenase 2
nDNA - Nuclear DNA
ng - Nanogram
PCoA - Principal coordinates analysis
PCR - Polymerase chain reaction
s- Second
Trp - Tryptophan
vs - Versus
yr - Year
ΔK - Best number of clusters estimated by STRUCTURE

µl - Microliter

µM - Micromolar

16S - Ribosomal RNA

16sH - Heavy strand of 16S primer (Rv)

16sL - Light strand of 16S primer (Fw)

1st intron of S7 - Ribosomal protein gene

1. Introduction

1.1 Biodiversity patterns and processes

Understanding the factors that generate and maintain biological diversity are fundamental pursuits in evolutionary and conservation biology. By identifying these factors we can better understand evolutionary processes such as ecological adaptation and speciation (Pimm 2008). Geographic patterns in organismal diversity can be shaped by both biotic and abiotic factors, which often interact and can operate on various temporal and spatial scales (Gaston 2000). Therefore, a multiple-disciplinary research approach, including the use of the genetic and geographic information, is often necessary to understand how species evolve and diversify.

An important step in understanding the factors that shape patterns of biological diversity is to identify an appropriate way in which to measure diversity itself. One of the simplest measures of diversity is species richness. Species richness is simply a count of the number of species in a given area without correcting for abundance of other factors (Gotelli & Colwell 2001). However, it is often difficult to identify species boundaries and measures of species richness do not take into account intraspecific variation. An alternative method for measuring biological variation is to measure “biodiversity”. The term biodiversity was first introduced by Lovejoy (1980) to describe the number of species in a given area, or species richness. Since then the definition of biodiversity has changed to accommodate genetic and morphological variation within species, as well as community-level processes (Swingland 2000). For example DeLong (1996) states that : ”Biodiversity is an attribute of an area and specifically refers to the variety within and among living organisms, assemblages of living organisms, biotic communities, and biotic processes, whether naturally occurring or modified by humans.” Biodiversity can be measured in units of genetic diversity, which helps to identify species boundaries and identifies significant variation within species. Measuring biodiversity in units of genetic diversity can be especially helpful for conservation purposes, as it may aid in the identification of divergent lineages within similar morphological characteristics (i.e., cryptic species; Dickman *et al.* 2007). Focusing on genetic diversity rather than species themselves can help to identify the factors that shape patterns of biodiversity in different regions and may promote speciation.

Geographic patterns of biodiversity can be influenced by physical factors. For example, plate tectonics and climate changes, including oscillations of temperatures, glaciations and changes in the sea level are all thought to have influenced patterns of global biodiversity (Brown & Lomolino 1998). Oscillations in environment factors are important because they can affect the distribution of species and their capacity to disperse (Dynesius & Jansson 2000). For example, during and after the Pleistocene, vicariant events such as habitat fragmentation due to climatic changes and changes in sea levels are thought to have occurred for many species that currently inhabit temperate areas. Such vicariant events often resulted in isolation of populations and species and divergence via allopatric processes (Brown & Lomolino 1998).

Biological factors such as the dispersal ability of an organism can also be important for shaping patterns of biodiversity. The ability or propensity to disperse often differs within and among species and can have profound effect on the rates at which they diverge. Greater dispersal can increase gene flow among populations and constrain evolutionary divergence (Bohonak 1999). Alternatively, limited dispersal may restrict gene flow within a species and can promote population divergence via neutral or adaptive processes such as genetic drift or ecological adaptation. Species with limited dispersal abilities often exhibit stronger genetic differences among populations and within them (Blouin *et al.* 2010). Reduced gene flow may allow populations within a species to diverge in response to spatially varying selection pressures and increase rates of local adaptation (Kawecki & Ebert 2004).

A single climatic event may affect species living in the same region, but in different habitats (Collin & de Maintenon 2002), in unique ways due to specific interaction of biotic and abiotic factors. Vicariant events split continuous populations or species into multiple parts via geologic barriers that prevent gene flow among them. For example, climatic changes may make areas within the range of a species unsuitable for them to inhabit and isolate its populations in discrete habitat patches. Subsequent environmental changes may permit dispersal across such barriers and allow species to successfully colonize new areas (Avise 2000). The frequency and severity of an abiotic change that affects vicariance and dispersal may shape patterns of biodiversity within different groups of species in dissimilar ways. For example, organisms with reduced geographical range, high habitat specificity, or lower dispersal capacity may be more affected by vicariant events such as habitat loss and destruction (Frankham 2005).

1.2 Desert environment

Species living in deserts, which fully depend on humid habitats and dispersal corridors, are likely to be negatively impacted by climatic changes and habitat fragmentation. Deserts are among the harshest ecosystems on Earth, characterized by very low precipitation, highly fluctuating temperatures and limited resources (Ezcurra *et al.* 2006). Organisms adapted to arid and semi-arid conditions are often highly specialized to inhabit the harsh desert climate (Ezcurra *et al.* 2006; Ward 2009). Climate variation and human activity can often cause a desertification effect and each desert and (or) area of the desert can be affected differently according to their characteristics. Desertification often increases fragmentation in arid habitats with standing water. Species adapted to survive in what little water exists in desert environments may be at risk from climate change and increased habitat fragmentation (Ward 2009). Limited dispersal due to climatic change induced habitat fragmentation may be especially detrimental to species in inhabiting desert environments.

In West Africa, climate fluctuations over the past seven million years may have had a major impact on rates of species diversification and overall levels of biodiversity in the region (Ward 2009). The Sahara is the largest and driest desert in the world, which is bordering with the semi-arid region, Sahel, in West Africa. Sahel characterized by variation of a few rainy seasons and the dryness for the rest of the year (Nicholson 1995; Ward 2009). The onset of desert-like conditions in the area that is now referred to as the Sahara started about seven million years ago (Ward 2009), although in some areas, like Mauritania, it is estimated to be slightly younger (around 6 to 2.5 Mya). Since the Pliocene (5.3 to 2.5 Mya), the Sahara-Sahel has fluctuated between wet and dry climate (Figure 1) (Le Houérou 1997). The area of Sahara was much larger and climate much warmer at the Last Glacial Maximum (LGM; 18,000 yr) (Holmes 2008). The period when the Sahara was covered by vegetation, lakes and wetlands is called Green Sahara which occurred 7,000 years ago at the mid-Holocene and ended about 6,000 years ago, when humid climate was replaced by arid climates (Holmes 2008; Kröpelin *et al.* 2008). During the drier periods, many aquatic species were likely isolated in smaller residual lakes and ponds in the Sahara. As rivers formed and ponds and lakes grew in size during wetter periods, connectivity among populations probably increased. For example, periods of isolation may have facilitated intraspecific divergence and promoted speciation (Gonçalves *et al.* 2012). At other times, gene flow may have increased

genetic diversity in isolated and potentially small and genetically depleted populations. The relatively high number of endemic species found in the Sahara-Sahel may be due to the effect of such climate fluctuations on connectivity (Brito *et al.* 2014; Ezcurra *et al.* 2006).

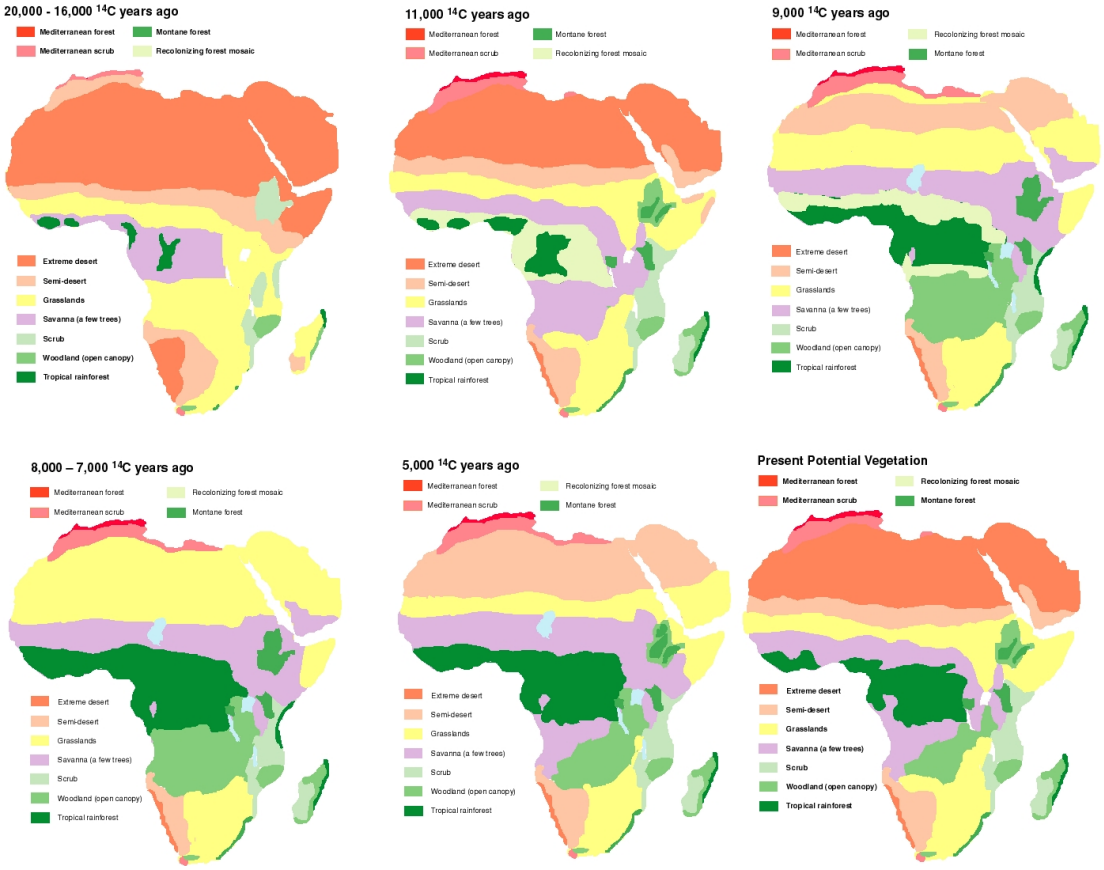


Figure 1. The Sahara climate fluctuations. Representation of climate fluctuation in Sahara since the Last Glacial Maximum until present time; adapted from (Adams & Faure 2004)**.**

Recent studies suggested high rates of endemism and the occurrence of "micro-hotspots" of speciation throughout the Sahara (Wagner *et al.* 2011). Compared to nearby areas such as Eurasia, little is known about biodiversity in the Sahara-Sahel, as it is generally a more difficult area to travel in and research there has been hampered by greater political instability. However, studies suggest the Sahara-Sahel suffered a catastrophic decline not only in hotspots but also in mega fauna, probably due to human alterations to sensitive habitats. Many species were adapted to harsh and highly variable desert environment (Brito *et al.* 2014; Durant *et al.* 2014). The mountains of Adrar Atar, Tagant, Afollé and Assaba in Mauritania, surrounded by the harsh Sahara desert, comprise isolated populations of several organisms, including mammals (Vale *et al.* 2012a), fishes (Trape 2009, 2013) and

reptiles (crocodiles (Brito *et al.* 2011; Velo-Antón *et al.* 2014), lizards (Vale *et al.* 2012b). The hotspots in the mountains (Brito *et al.* 2011; Trape 2009) are still under studied at phylogenetic, phylogeographic, and population genetic levels, where rich biological assembles are present (Brito *et al.* 2014).

1.3 Water availability in Mauritania

Changes in the geographic distribution of water resources in the Sahara likely had a major influence on how biodiversity has evolved in aquatic organisms in the region. Understanding the factors that have shaped patterns of gene flow within species among contemporary water bodies in the Sahara-Sahel will help to understand how past dispersal events may have affected rates of diversification. For many aquatic organisms the dispersal ability can be affected by their different habitats, which can be permanent, seasonally present due to rainfall or present only by flood events (ephemeral) (Fagan 2002). Ephemeral habitat may be important corridors for dispersal in arid and semi-arid regions, providing opportunities for connectivity (Haas *et al.* 2009). Different types of water resources can be identified by different levels of vegetation cover, inhabiting fauna and aerial measures of turbidity (Lacaux *et al.* 2007). Quantifying number, size, and distribution of water-bodies in desert habitats are crucial for understanding connectivity patterns.

The mountains of Mauritania are generally surrounded by two types of water-bodies: *gueltas* and *tâmoûrts*. *Gueltas* are rock pools ranging between 0.001 ha and 1.0 ha, located upstream of valleys at the base of mountains (Brito *et al.* 2011; Cooper *et al.* 2006). *Tâmoûrts* are seasonal wetlands, mostly arid during dry season, but during the wet season are normally large (more than 1000 ha), located on the foothills of mountains (Cooper *et al.* 2006). These water-bodies can be either permanent or seasonal. When they dry out local extinction of endemic aquatic species may occur if species are not able to find micro-habitats (to aestivate) or if corridors – connections to other areas – are lacking (Julian D. Olden *et al.* 2010). The existence of past water corridors may have decreased the extinction risk of some species by increasing dispersal rates between isolated populations (Simberloff *et al.* 1992). Corridors suitable for dispersal of fish species among water bodies are formed along raging streams. These temporary corridors usually flow to vast plains adjacent to the *gueltas*, during the rainy season (July to September). These

periodical and short time events can connect some of the isolated habitats for a limited period of time.

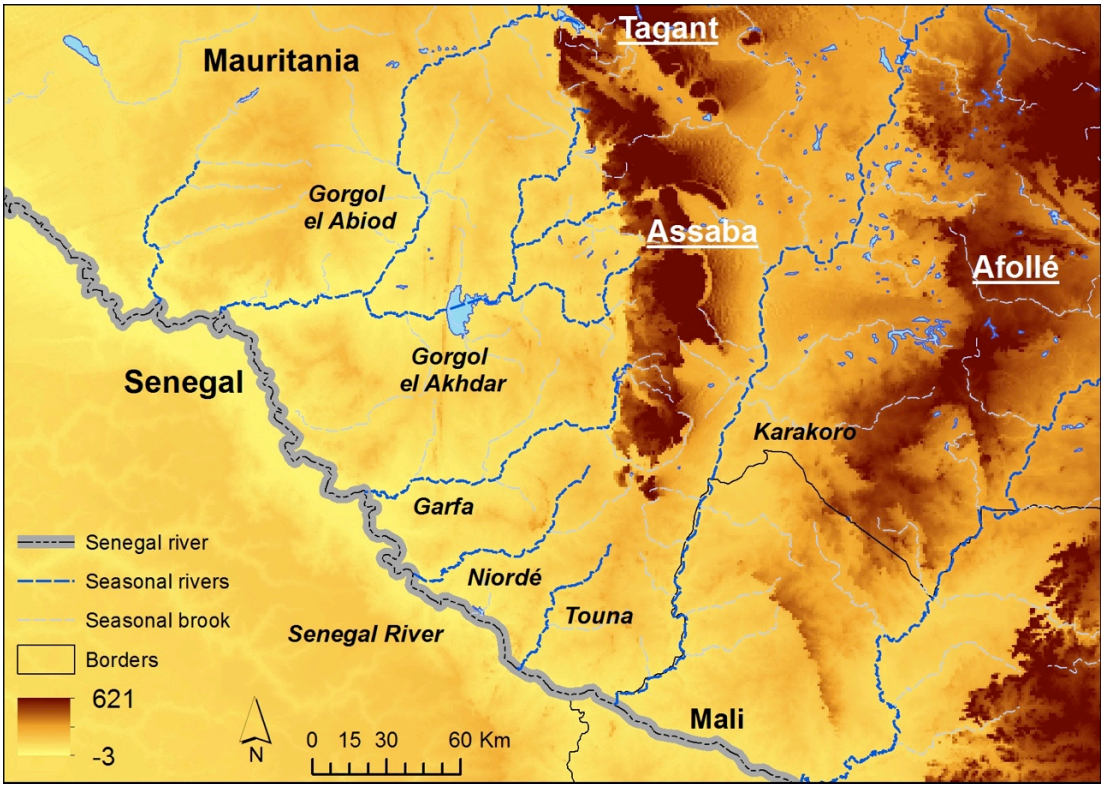


Figure 2. Main mountains and rivers of southern Mauritania. The names of mountains are underlined, the major seasonal rivers and Senegal River in italic letter**.**

The southern mountains of Mauritania are connected to the Senegal river through seasonal rivers (Figure 2) (Cooper *et al.* 2006), which may allow aquatic organisms to intermittently disperse between them. Analyses of satellite imagery suggested that Gorgol el Abiod, Gorgol el Akhar and Garfa sub-basins in the Assaba mountain, assure temporary connections (Figure 3) between some isolated *gueltas* and also between *gueltas* and Senegal river basin (Campos *et al.* 2012). In Brito *et al.* (2011) it was shown that crocodiles (*Crocodylus suchus*) were dispersed among mountain *gueltas*, and also between *gueltas* and the Senegal river via these temporary rivers. The authors suggested that these temporary corridors for dispersal might be important in shaping patterns of genetic diversity within aquatic organisms in the region.

Changes in water availability also appear to be an important determinant of dispersal and gene flow in aquatic organism in the Karakoro river basin. Water availability in Karakoro river basin near Afollé mountain in the eastern part of Mauritania is generally more limited than in the western areas near Assaba mountain

(Campos *et al.* 2012), probably due to the geomorphological characteristics of the area where the bend plateau of the mountain declines to the west (Toupet 1966). Aridity in the region, coupled with human activities that further deplete water sources, may form barriers to dispersal for several mammal species in the Karakoro basin (Vale 2010). Connectivity via these seasonal rivers may be disrupted by climate change and human disturbance such as water extraction for agriculture, domestic and cattle needs (Brito *et al.* 2014) and have significant impact on biota in the area.

Figure 3. Seasonal rivers. Two rivers during the dry season flowing from the Assaba mountain (at the right) and joining together (at the left), and connecting with Senegal river, about 150 km the southwest. Photo: Zbyszek Boratyński**.**

In this area, there is still a lack of knowledge concerning the distribution and dispersal patterns especially of small size animals, such as amphibians and fishes (Brito *et al.* 2014). Organisms, such as fishes, may present a good model to comprehend the hydrological connection between isolated water-bodies and basin due to their dependency on water. More specifically, the relict Tilapia fish populations may present a suitable biological model for studying the evolutionary processes and phylogeographical patterns in these particular areas due to their ability to adapt to different environments.

1.4 African Tilapias

Tilapias (*Tilapia* Smith, 1840 and related species) are members of Cichlidea (cichlids) family, which is known as the most species rich family of vertebrates, including more than 3.000 species (Kocher 2004). Cichlids have a broad distribution in Central and South America, through Africa, Madagascar and southern India (Chakrabarty 2004) in coastal, brackish and freshwaters (Beveridge & McAndrew 2000). The two great tribes, Tilapiini and Haplochromini, are broadly distributed in Africa. Tilapia is common name for many cichlid species (from the Tilapiini tribe), but especially for species of three genera: *Coptodon (*known before as *Tilapia), Sarotherodon* and *Oreochromis.* The last two genera are biparental/paternal and maternal mouthbreeders, respectively (Trewavas 1983). Tilapias hold complex systematics and taxonomy, attracting many scientists. Together with other cichlids they show complex behavior, great example of adaptive radiation (especial from East Africa cichlids) (Beveridge & McAndrew 2000; Kocher 2004) and are important for the aquaculture (Beveridge & McAndrew 2000). The East African cichlids endemic to Great Lakes (Tanganyika, Malawi and Victoria) were studied at morphological, behavioural and ecological diversity levels which lead cichlids to their evolutionary successes (Salzburger *et al.* 2005).

Many studies were carried out to improve the knowledge of the number of genera within Tilapias and the phylogenetic relation among them, using allozyme approaches with morphological or behavioural studies (Kornfield *et al.* 1979; McAndrew & Majumdar 1984; Sodsuk & McAndrew 1991) and later including the DNA sequencing approaches (Franck *et al.* 1992; Franck *et al.* 1994; Sültmann & Mayer 1997; Sultmann *et al.* 1995).

Schwarzer *et al.* (2009) performed, for the first time, phylogenetic analyses of African Tilapias based on the multilocus DNA data (from nine mitochondrial and nuclear sequences of nine markers) with geographic distribution of each clades. They revealed Tilapias are paraphyletic groups, and the term “Tilapiini”, commonly used in previous studies, is incorrect in the phylogenetic context. Also, they described three new groups, one of which is Oreochromini tribe (the mouthbreeding genera) that includes the *Sarotherodon* and *Oreochromis* genera. Since earlier studies they are mentioned as very close genera. Moreover, the most recent work of Dunz and Schliewen (2013), using the information from the previous studies including evidence of morphology, proposed a novel classification of 20 haplotilapiine cichlid genera and nine tribes (Figure 4). The clade with *Tilapia* species was renamed to *Coptodon* species and the tribe was called as Coptodonini,

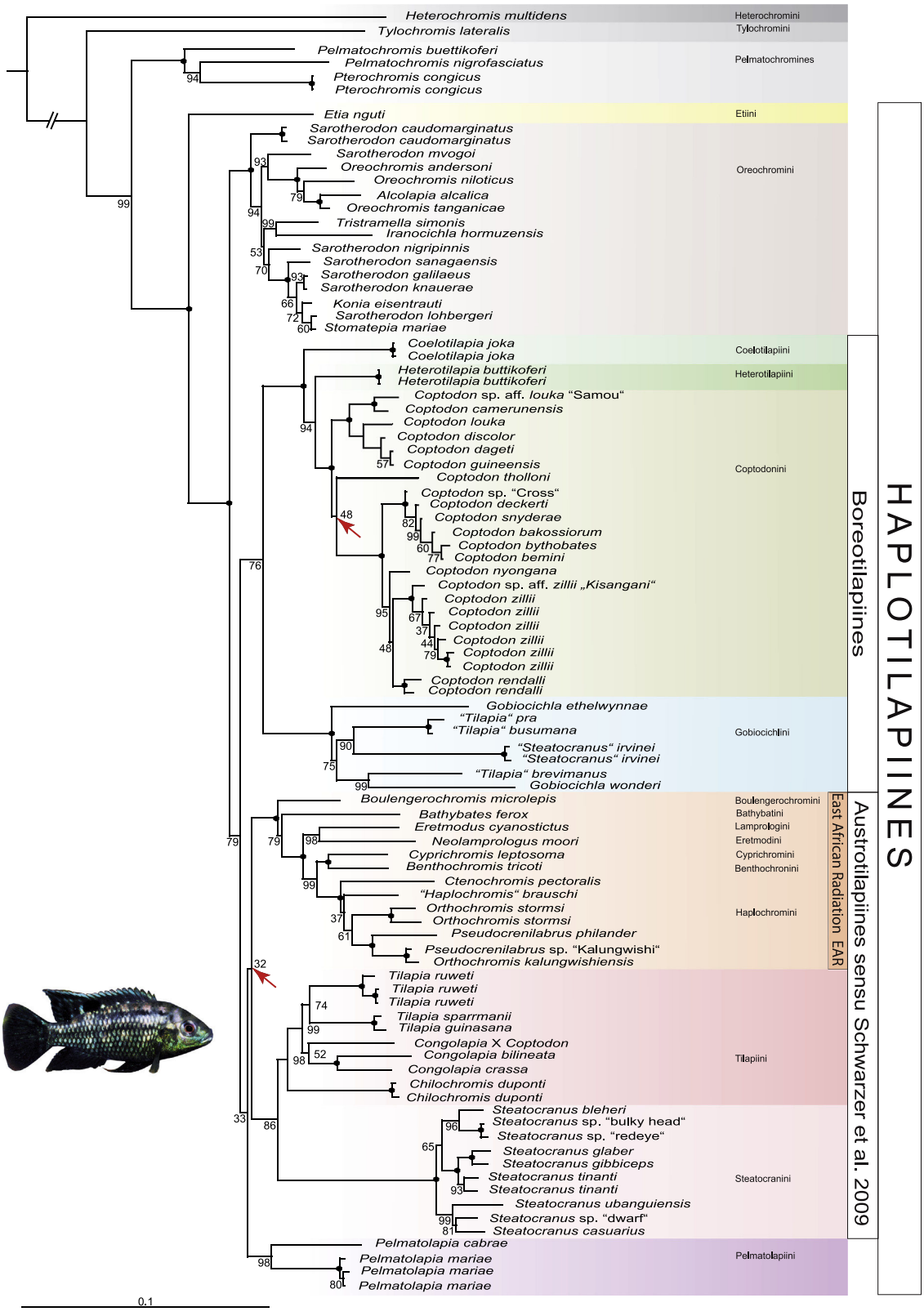


Figure 4. Phylogenetic relationship of Haplotilapiines (94 taxa), reported by Dunz & Schliewen (2013). This analyses were based on the consensus data of four mitochondrial DNA and five nuclear DNA g. The nodes of BS 100 (ML) marked by black hexagon and by red arrows the nodes, which differ in the BI and ML analyses

while before *Coptodon* was known as subspecies of *Tilapia*. These studies give a better idea about phylogenetic relation among genera/species and highlight their complexity regarding systematics and taxonomy. Nevertheless, there are still non-sampled cichlids in Africa, especially in North-West Africa, that may improve the knowledge of taxonomy and phylogenetic studies in African Tilapias.

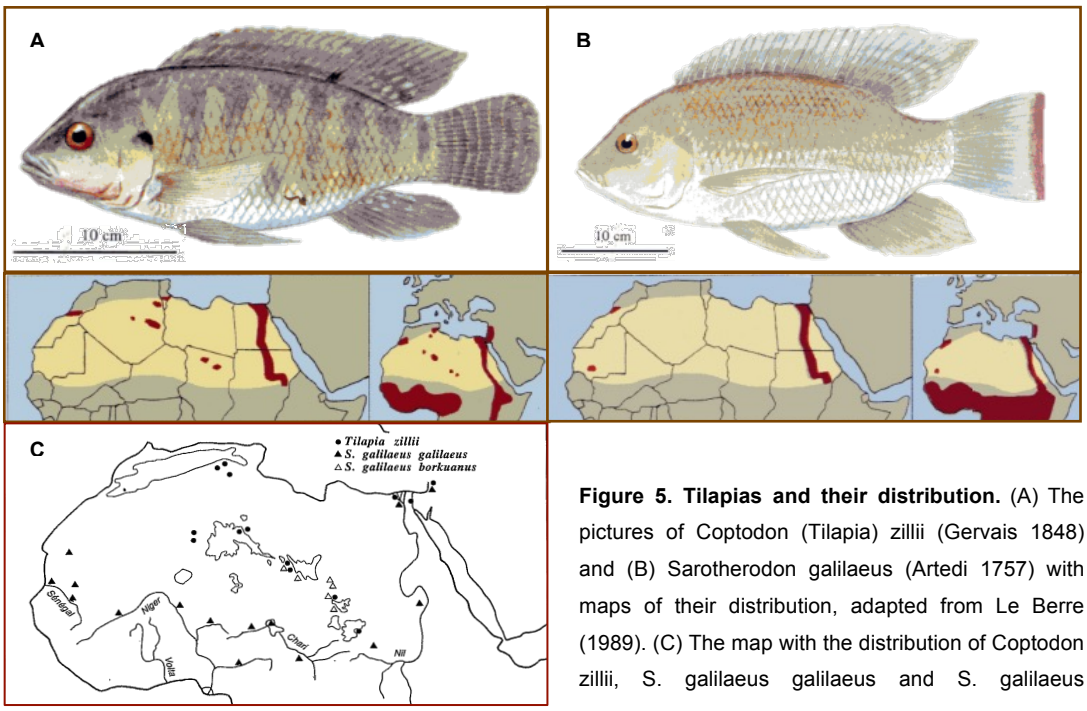


Figure 5. Tilapias and their distribution. (A) The pictures of Coptodon (Tilapia) zillii (Gervais 1848) and (B) Sarotherodon galilaeus (Artedi 1757) with maps of their distribution, adapted from Le Berre (1989). (C) The map with the distribution of Coptodon zillii, S. galilaeus galilaeus and S. galilaeus borkuanus (Lévêque 1990).

Many fish fauna in the Sahara, which were first recognized during expeditions to Africa in the earlier 20th century (Dekeyser & Villiers 1956; Le Berre 1989; Monod 1951; Villiers 1953), may have survived for thousands of years in isolated populations in water-bodies located at mountain massifs. Lévêque (1990) updated the taxonomic list of these relict tropical species and their distribution maps, with some remarkably broad distributions, and illustrated the presence of two species of cichlids in North and West Africa (*Sarotherodon* and *Coptodon (Tilapia)*) (Figure 5). Due to a short time period of isolation, speciation processes and endemism are unexpected, according to the author. However, Trape (2009) focused on studying the distribution of the recorded relict populations found some populations of Adrar (Mauritania) were extinct and recorded two new species from the northern Chad (Trape 2013). These studies highlight that our knowledge about the distribution of relict fauna is largely incomplete and old distribution maps (Figure 5) are clearly coarse. More studies on non-sampled areas and molecular approaches would improve our knowledge about distribution and evolutionary processes of fish taxa in

Sahara. Tilapias as other relict fishes are located in restricted water bodies in which they survived since the beginning of the dry period (Lévêque 1990), and temporal rivers can potentially act as corridors between closely/distinct isolated habitats or connect these to basins, during the wet seasons.

1.5 Molecular markers to address biodiversity patterns and processes

The biodiversity patterns of species are a result of evolutionary processes leading to their existence and of their disability to disperse due to different barriers, (Avise 2000; Tokeshi 2009). The change of individual traits, as morphology, life history, physiology and behaviuor, are usually affected by the interaction of various genes and the environment (Allendorf & Luikart 2007). Genetic diversity within populations and genetic differentiation among populations are measurements commonly used (Allendorf & Luikart 2007), which can be explained by habitat fragmentation, climate changes or other dynamics in ecosystems. Neutral genetic variation can help to identify levels of structure, gene flow and extinction risk in biodiversity ecosystems (Etienne & Olff 2004).

Molecular tools with different sensitivities to historical and demographic processes as mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) at different locus are good markers for phylogenetic and population genetic studies. Since mtDNA has higher mutation rates comparing with nDNA (Brown *et al.* 1979), and is nonrecombinant and maternally inherited, it can assist the identification of gene flow during the recent past divergence processes (Marsjan & Oldenbroek 2007). On another hand, microsatellite markers (nDNA) are commonly used for studying more recent molecular events in populations. Microsatellites are small fragments of DNA with repeat units in sequences located in non-coding nDNA (Marsjan & Oldenbroek 2007). These markers exhibit high polymorphism, are sensitive to changes in effective population size and the processes as genetic drift and mutations can make rapid changes in allelic frequency (Cañón *et al.* 2001).

It is important to use multiple, unlinked, neutral genetic loci to estimate divergence times between lineages or species. Regions of the nuclear genome may exhibit different rates of evolution, gene loss and duplication, and recombination (Mitchell & Gonder 2013). Consequently, estimates of divergence time between populations or species may vary due to some loci having different evolutionary

histories and coalescence times. Divergence times based on multi-locus data may be more accurate as they can be calculated using methods that correct for the different evolutionary histories of loci (Edwards & Beerli 2000), including those from both nDNA and mtDNA genomes of species. More accurate inferences about the evolutionary histories of different taxa may be possible using multi-locus datasets.

1.6 Objectives

The main objectives of this thesis are to describe population genetic structure in Tilapia fishes from Mauritania and attempt to identify barriers to dispersal and gene flow within them. This study may help to understand the evolutionary history of Tilapias in the Sahara, including the role that post-Pleistocene corridors for dispersal have shaped patterns of genetic diversity within and among species or subspecies in West Africa.

In order to achieve the main goal of the study, we performed phylogenetic analyses to identify taxonomic groups present in Mauritanian mountains and water bodies, to clarify at which level further genetic population studies will be performed (at intra- or interspecific level). Also, including Tilapia specimens collected across North-West Africa we can get a better idea about the evolutionary processes of these species. The phylogenetic analyses were performed using two mtDNA loci and the nDNA fragment, chosen from the multi-locus data used in phylogenetic studies of other African cichlids. We developed specific microsatellite loci for the taxa present in Mauritania inland, to accomplish population genetic analyses. The results were combined with the information obtained from the sequencing data. Five main objectives were identified to answer the following five questions, at phylogenetic and population levels:

1) What is the phylogenetic relationship among the Tilapia samples collected from North-West Africa and their relationship within African cichlids?
2) What is the distribution of the obtained lineages?
3) Are mountain populations isolated from lowland populations?
4) How hydraulic network are related to gene flow?
5) How is the genetic diversity distributed among the genetic groups?

2. Material and Methods

2.1 Study area and samples

This study is primarily focused on individuals collected from different areas from Morocco to Mauritania, including coastal and inland sites (Figure 6). A few specimens from Niger were also examined. The samples from Mauritania were collected from two mountains: Assaba and Afollé, and five seasonal sub-basins: Rkiz, Gorgol el Akhdar, Garfa, Niordé and Karakoro, represented by seasonal and permanent mountain water bodies (*tâmoûrts* and *gueltas*) connecting to the Senegal River (Campos *et al.* 2012). Connectivity between mountain and lowland areas in Mauritania occurs during the rainy seasons from July to October, with the highest precipitation occurring in August and September. The rest of the year, this area is dry and cool (November to January) and hot (March to June) (Cooper *et al.* 2006).

Tilapia samples were collected during multiple field trips to Africa, from 2007 to 2014. One hundred and ninety-nine samples from 48 locations, 15 sub-basins and 4 basins in the North-West Africa were used in this study (Table S1, **Appendix**). Samples included 174 fresh and 25 near-fresh tissues (from specimens found dead). Whole specimens or tissue samples were stored in 95% ethanol and maintained at room temperature.

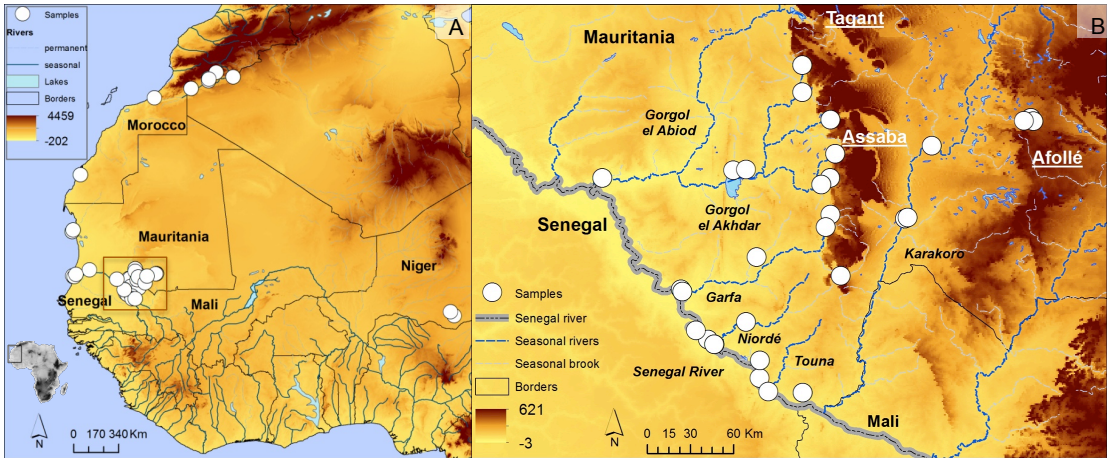


Figure 6. Study area. (A) Geographic distribution of samples collected in Morocco, Mauritania, Senegal and Niger, to be used in phylogenetic analyses. (B) The small inset map presents the main study area – Mauritania. Each dot is the coordinate where sample were collected, one dot can have more then one samples.

2.2 DNA extractions and sequencing

Fish samples, from fin or muscle tissue, were extracted using the EasySpin® Genomic DNA Tissue Kit (Citomed) or the QIAamp® DNA Micro Kit (Qiagen), following the manufacture's instructions. Some DNA extractions were diluted with ddH_2O, to prevent inhibitors from interfering with the PCR reaction. Two mitochondrial and one nuclear fragments were PCR amplified: 16S (500bp; primers 16SL and 16SH from Hillis *et al.* 1996), partial encoding ND2 gene (735bp; primers ND2Met and ND2Trp from Klett & Meyer 2002) and first intron of the S7 ribosomal protein gene (466bp; primers S7RPEX1F and S7RPEX2R from Chow & Hazama 1998).

PCR reactions for the mtDNA and S7 loci were performed in a final volume of 10 µl, containing 5µl of MyTaq™ HS Mix (BioLine), 0.4 µM of each primer, 3.2µl ddH_2O and approximately 10ng of genomic DNA. Both mtDNA genes were amplified under the following PCR conditions: an initial denaturation at 95 °C for 10 min, followed by 40 cycles of denaturation at 95 °C (30 s), annealing at 62 °C (45 s), extension at 72 °C (45 s) and a final extension for 10 min at 72 °C. A touchdown PCR cycling program was used to amplify S7 fragment, where annealing temperature were decrease 0.5 °C each cycle from 62 °C to 58 °C, totaling 40 cycles. The PCR programs were modified in some cases for samples with low quantities of DNA. Specifically, the annealing temperature was decreased to 58 - 56 °C. The re-amplification of PCR products was used in some cases.

PCR products were purified using ExoSAP-IT® PCR clean-up Kit (GE Healthcare) and sequenced following the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) protocol in one direction. Sequencing products were electrophoresed on a 3130*xl* Genetic Analyzer (Applied Biosystems). Sequences of each gene were aligned and edited in software SeqScape v2.0 (Applied Biosystems). Any resulting DNA sequences with ambiguous bases were sequenced in the reverse direction and re-edit. The resulting mtDNA and 1st intron of S7 data was aligned together with 52 sequences (Table S2, **Appendix**), chosen among the previously published in Dunz and Schliewen (2013), using Geneious v4.8.5 software (Drummond *et al.* 2009).

2.3 Microsatellite development, amplification and genotyping

Microsatellites loci were developed for the *Sarotherodon* mitochondrial B lineage (see results for details), using thirteen individuals collected in different locations in Mauritania. Genomic DNA was extracted using EasySpin® Genomic DNA Tissue Kit (Citomed). The DNA in each sample was quantified in fluorometer, using Quant-iT™ PicoGreen®dsDNA Assay Kit (Invitrogen). Equimolar pooling of the 13 samples totaling 3 µg of genomic DNA were sent to Genoscreen, France (www.genoscreen.fr), for microsatellite development at the 454 GS-FLX Titanium pyrosequencing platform (GS FLX®, Roche Diagnostics) through enriched DNA libraries (Malausa *et al.* 2011). Total DNA was enriched for 32 different repeat motifs. Briefly, GS-FLX libraries were constructed following manufacturer's protocols (Roche Diagnostics) and sequenced on a GsFLX-PTP. The bioinformatics program QDD (Meglécz *et al.* 2010) was used to filter for redundancy, resulting in a final set of 1549 sequences from which 261 primers pairs were designed. Thirty primer pairs were chosen for testing (see for details in Primer Note, **Appendix**), out of which thirteen (Table 1) were used for further population genetic analyses.

A total of 132 individuals from mitochondrial lineages *Sarotherodon* A (n=12) and B, (n=120) were screened at 13 microsatellite loci. Loci were combined into two multiplex and one singleplex. Amplifications of multiplexes were performed in a final volume of 10 µl, containing 5µl of Qiagen© Multiplex PCR Kit Master Mix (Qiagen), 1 µM of primer mix (Table 1), 3µl ddH_2O and approximately 10 ng of gDNA. PCR cycles consisted of an initial denaturation at 95 °C for 15 min, following by 9 cycles of denaturation at 95 °C (30 s), annealing at 56 °C (1 min 30 s) with decrees in each following cycle by 0.5 °C, extension at 72 °C (45 s); 31 cycles at 95 °C (30 s), annealing at 52 °C (1 min), extension at 72 °C (45 s) and final extension step at 60 °C (30 min). For some samples that did not amplify, the multiplex was splitted into halves or even until singleplex. For singleplex reactions primers were used in the following concentrations: 0.04 µM of forward primer, 0.4 µM of reverser primer and 0.4 µM of a M13-tailed fluorescent labeled primer (Table 1; (Blacket *et al.* 2012)). Similar thermocycling conditions were used but final extension was shorted to 10 min. PCR products were visualized in a 2% agarose gel and fragments were separated by size on a 3130*xl* Genetic Analyzer (AB), using GeneScan™ 500 LIZ® Size Standard (Applied Biosystems). Allele scores were determined using GeneMapper v4.1 (Applied Biosystems) and checked manually. Samples extracted

from near-fresh tissue were amplified at least for three PCR replicates to minimize amplification errors as allelic dropouts and false alleles.

Table 1. List of 13 microsatellites developed in *Sarotherodon* mtDNA lineage B along with locus name, multiplex (Mix), repeat motif, fluorescent label M13-tail used, primer sequences, primer concentration (μM), allele size range (bp).

Locus	Repeat motif	Fluor. label	Primer sequences (5´-3´)	Primer (μM)	Size range (bp)
Til6	$(AC)_{13}$	VIC	F: CCCTCCTGCTTAAGTCAACAC R: TAAGGGTTCCCGACTCTCCT	0.4	115 - 151
Mix1					
Tlil1	$(ATAG)_{12}$	6-FAM	F: TGCTAAGAACAAAGTCTTGGGA R: TCATCATGCTGCGGTAACAC	0.4	222 - 266
Til2	$(CA)_{14}$	6-FAM	F: TGGAACACTTTGGAGGATCG R: CGTTCCTGGATCAAACCG	0.4	178 - 220
Til13	$(CCAT)_{10}$	6-FAM	F: GAGCTTGGCCATTTAGGAGA R: TGTTGGAACATAAGGATAAGTGGA	0.8	120 - 188
Til8	$(CAA)_{15}$	NED	F: CCCACAAAAGAATTAAAGCTAAAGA R: TGAGTCTAACATTTGGTCTTTGAA	1.7	110 - 158
Til19	$(TCTA)_{15}$	PET	F: ATCAGGGTCGTGACTTCTGC R: CCAAATGAGCAGGATGATTG	1	164 - 268
Til20	$(CA)_{13}$	PET	F: CAGCTTCCCGTGCTCATTAT R: GCAGCACCTGGAAGATAAAA	1.8	130 - 148
Mix2					
Til21	$(TGGA)_{9}$	6-FAM	F: GGTCCAAAGACATAATTAGTGGG R: TGTGTTGGGTAAGTATCTTCTAGTTCA	0.4	260 - 298
Til22	$(GT)_{12}$	6-FAM	F: AAACCAATAAATACAACCCACTG R: TTAACACAGTTTCTTATACGACAGAGA	0.4	154 - 182
Til23	$(CA)_{10}$	6-FAM	F: TTAGTCAGCAGCCAACACCC R: GACCTCCTCGTGAAGTGCAT	0.35	104 - 114
Til24	$(AC)_{8}$	VIC	F: TGAACGTGCAGCAGAGTCAC R: GGAGCTCTTTAGGTGGTTGG	0.4	230 - 234
Til25	$(CA)_{14}$	VIC	F: GGGGCCTCAGTCTAGGAAAG R: GGAGTGGGGACGTGCATA	0.5	164 - 216
Til30	$(TG)_{11}$	PET	F: CTGACAGCAAGAGCCTCAAA R: GAAATAGGTTCTGGCTCAAAGTT	0.6	111 - 121

2.4 mtDNA and nDNA analyses

One hundred ninety-one samples were successfully amplified for both mitochondrial DNA markers. Forty-seven samples comprising at least one individual per location were selected to be PCR amplified for the nuclear intron. Partitioned Bayesian analyses were used to perform two phylogenetic trees, based on haplotypes of mtDNA (1239bp, 16S+ND2) and nDNA (492bp, 1st intron of S7) sequences, respectively. nDNA sequences were phased, and haplotypes of both genomic compartments were estimated using software DnaSP v5.0 (Rozas & Rozas 1995). 10.64% of individuals with ambiguous genotypes remained unresolved with the confidence probability interval lower than 0.9, and those sites were marked as missing data (N). Recombination can mislead phylogenetic analyses in several ways leading to the overestimation of the number of mutations within a dataset and the

underestimation of the time to the most common recent ancestor (Schierup & Hein 2000). We therefore checked the 1st intron of S7 haplotype data for evidence of recombination using DnaSP v5.0 software. The three codon positions of ND2 marker were approached as a separate data partition. The best partitions and molecular evolutionary models for each partition was identified using Partition Finder v1.1.1 (Lanfear *et al.* 2012) (Table S5, **Appendix**). Bayesian analyses of the mtDNA genes with 4 partitions and separately of the nDNA gene were run in MrBayes v3.2.0 (Ronquist *et al.* 2012). Both trees were obtained from three parallel runs, using for each 5.0 x 10^7 million generations, with sampling every 1.0 x 10^3 generations. Convergence and stationary of runs were confirmed in Tracer v1.5 (Rambaut & Drummond 2007). The final phylogenetic trees were performed with 10% of burn-in. In order to construct haplotype networks and to estimate the relationship among them, we used the median-joining algorithm (MJ) (Bandelt *et al.* 1999) in NETWORK v.4.6.1.1. Undefined sites of the sequence alignments (16S+ND2 and 1st intron of S7) were masked, while insertion/deletions sites were coded by single nucleotide. Also haplotype network was performed on samples used for population genetics, based on both genes and represented in detail for each cluster.

2.5 Microsatellite data analyses

The multi-locus data was checked for errors and mismatches in Microsatellite toolkit (Park 2001). Two identical samples were discarded from further analyses. Consensus genotypes were built for individuals amplified for several replicates. Bayesian clustering and admixture analyses were performed in STRUCTURE v2.3.4 (Pritchard *et al.* 2000), using sub-basins as putative populations (Table S6, **Appendix**). Ten independent simulations were performed for each K (1 to 14), with 1.0 x 10^6 MCMC interactions and a burn-in of 4.0 x 10^5 interactions. The results were uploaded to the STRUCTURE HARVESTER 0.6.94 (Earl 2012) in order to identify the most likely number of genetic clusters (K) according to the highest value of ΔK (Evanno *et al.* 2005). A separate analyses was performed to identify the most likely number of genetic groups at each hierarchical level, applying the method of Evanno *et al.* (2005), until no further structure was found (Blouin *et al.* 2010; Phillipsen & Lytle 2013)

In order to visualize patterns of genetic differentiation among individual groups, a principal coordinate analyses (PCoA) was performed using GenALEx v6.5

(Peakall & Smouse 2006). Two different combinations of data were used: (1) with all individuals from both *Sarotherodon* A and B mtDNA lineages grouped by country and (2) within cluster including the majority individuals of Mauritania inland. Samples from each sub-basin were grouped by their topology in smaller groups, to examine the genetic structure pattern between lowland and mountains.

Genetic diversity was examined for each cluster identified in STRUCTURE and PCoA analyses for K=6. Three out of 128 samples were discarded from these analyses: two samples from a distinct area (Niger) and one sample that exhibited low posterior probability of assignment to a cluster (P<0.7) and thus could not be included in a single population. Deviation from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium analyses were performed in FSTAT v2.9.3.2 (Goudet 2001), for clusters with sample size > 6 (two clusters were excluded).

Mean number of alleles (N_a) per locus and number of private alleles was estimated for each of the six genetic clusters, while observed (H_O) and unbiased expected (H_E) heterozygosities was calculated for four clusters using GenALEx v6.5 (Peakall & Smouse 2006). The inbreeding coefficient (F_{IS}), allelic richness (A_R) per population and F_{ST} pairwise analyses between populations, were also calculated in FSTAT software. F_{ST} pairwise analyses were performed among the same four clusters with large sample size. Locus *Til20*, was excluded from this analyses because of high missing data exhibited for *Sarotherodon* mtDNA A lineage (Table S7, **Appendix**). Other comparisons within Mauritanian cluster were performed to test genetic differentiation among mountain and lowlands areas, although some interesting comparisons were not possible to perform due to small sample size, which could lead to over estimation of genetic differentiation or deviation from Hardy-Weinberg expectations (Willing *et al.* 2012).

3. Results

3.1 Phylogenetic mtDNA analyses

Thirty haplotypes (Table S2, **Appendix**) were found among 191 samples sequenced in this study for two mtDNA fragments (1239bp; 16S+ND2). Phylogenetic analyses using these sequences together with published ones (Dunz & Schliewen 2013) distinguished four mitochondrial lineages that were grouped in two different tribes Coptodonini and Oreochromini sensu Dunz and Schliewen (2013) and were named as *Coptodon* A and B and *Sarotherodon* A and B, respectively (Figure 7). These names are used for the entire study.

High phylogenetic support was found for both *Coptodon* A and B lineages within *Coptodon* species. *Coptodon* A (henceforth represented in blue) exhibited three haplotypes in 20 samples collected from coastal areas in Southern Morocco, in Banc d´Arguin National Park (PNBA, Mauritania) and a few samples from Senegal River basin in Mauritania, that were grouped with *Coptodon sp. aff. rheophila "Samou" (sp. aff.* unknown species which are close or similar) from Guinea and *Coptodon cameruensis* from Cameroon*,* with support values 0.94 and 1, respectively. *Coptodon* B (henceforth represented in green) presented five haplotypes in 38 samples collected from fresh waters in two sub-basins (Tassint and Zouwa) from Morocco and in three sub-basins (Touna, Bounoukolé and Karakoro), Senegal River from Mauritania (Figure 7). These haplotypes were grouped with all *Coptodon zillii* from Algeria, analyzed by Dunz and Schliewen (2013), with a high support value.

Sarotherodon A and B lineages are grouped with *Sarotherodon* genera species, but with different support values. *Sarotherodon* A (henceforth represented in red) evidenced four haplotypes in 13 samples, 11 of which collected in coastal areas in Banc d´Arguin National Park (PNBA, Mauritania) and two in Senegal basin close to Mauritanian coast, were grouped in a well-supported sister species relation with *Sarotherodon nigripinnis.* On the other hand, *Sarotherodon* B (henceforth represented in yellow), showed 18 haplotypes in 120 samples collected form Tissint, Draa and El Mellah sub-basins in Morocco, several *gueltas*, *tãmourts*, lakes and rivers near Assaba, Afollé mountains and Senegal River in Mauritania and also from Karoma sub-basin in Niger, showing a monophyletic relationship. Two Mauritanian haplotypes (the most distinct from other samples) exhibited a close relationship with

Sarotherodon knauerea from Cameroon and *Sarotherodon galilaeus* from Central-South Africa, with a support value of 0.72.

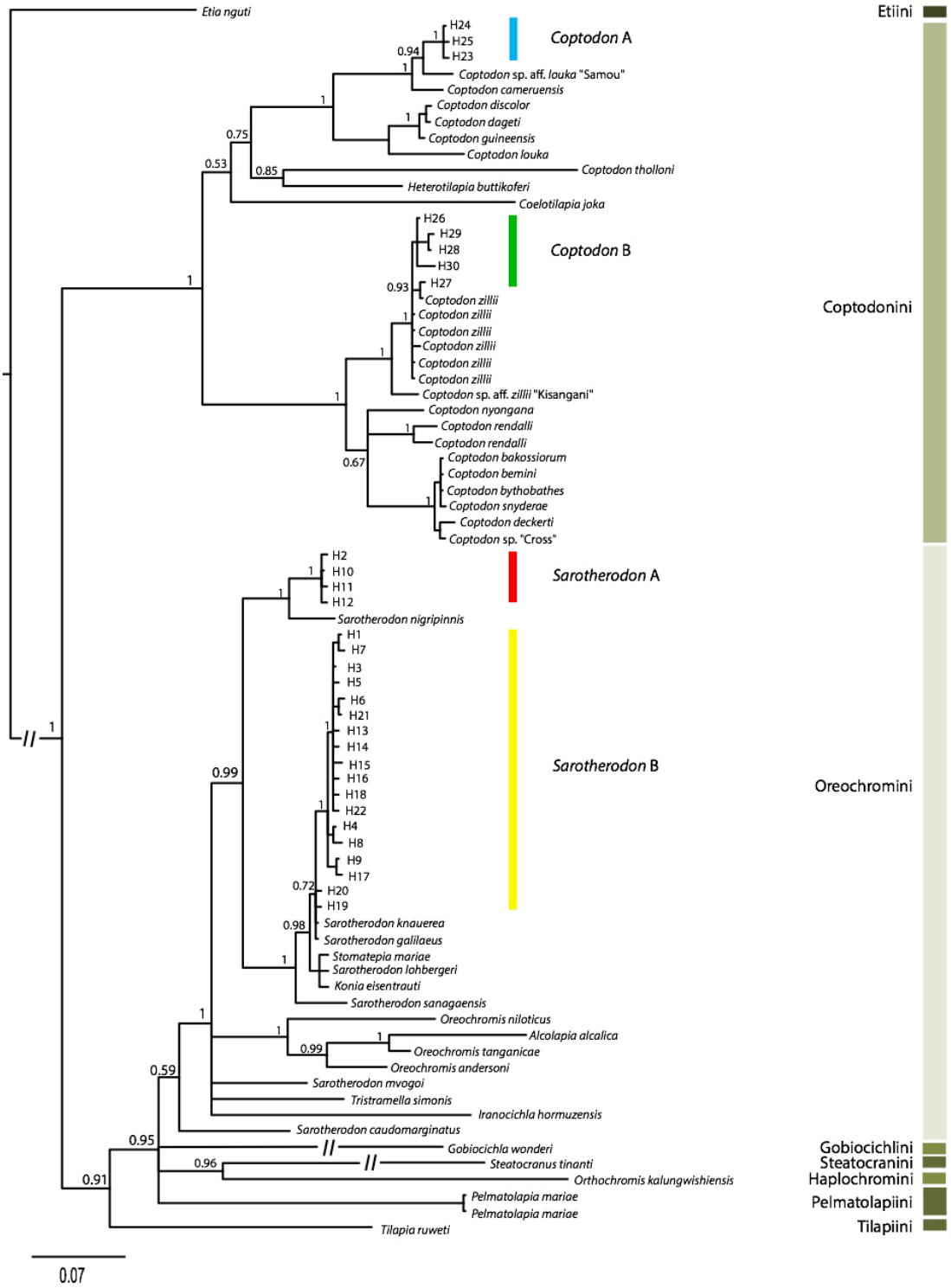


Figure 7. Phylogenetic relationship of four lineages based on two mtDNA markers (16S+ND2). A total of 30 haplotypes obtained in this study and 52 taxa (Table S3, Appendix) are presented. Bayesian analyses were performed in MrBayes. Each color represents a different lineage.

3.2 Recombination

Recombination was detected within two lineages at two separated locations within nDNA locus (1st intron of S7) (Figure S1, **Appendix**). One recombination event was detected in four haplotypes within the *Sarotherodon* B lineage that were found in one individual from Rkiz Lake (Rkiz sub-basin) and another individual from Bãfa (Gorgol el Akhdar sub-basin) in Mauritania. A second recombination event was detected between individuals that are part of two different lineages, *Sarotherodon* B and *Coptodon* B. Two haplotypes from *Coptodon* B lineage corresponding to samples from Senegal River and Karakoro sub-basin in Mauritania and the other two haplotypes from a lineage within several sub-basins in Morocco and Mauritania. Recombinant haplotypes were removed from all further analyses.

3.3 Phylogenetic nDNA analyses

Fourteen haplotypes (Table S4, **Appendix**) were observed on the analyses of 43 samples. Phylogenetic analyses based on the nDNA (492bp, 1st intron of S7) fragment resulted in a tree with similar topology for the two genera as the previously phylogenetic tree based on mtDNA. Nevertheless, within *Sarotherodon* and *Coptodon* we could not retrieve clades A and B as in mtDNA (Figure 8).

Main differences were obtained within each tribe. Three haplotypes from distinct *Coptodon* mtDNA lineages were grouped in one clade on nDNA, close to each other, and a fourth haplotype (H4) was more distinct. None of the nDNA haplotypes for *Coptodon* was grouped with the same species as in mtDNA phylogenetic tree. Two mtDNA lineages of *Sarotherodon* were grouped in the same nDNA clade and are represented in several paraphyletic sub-clades within this clade. We disclosed a new relationship between a *Sarotherodon* haplotype (H8) and *Oreochromis niloticus*. Three haplotypes of *Sarotherodon* (H9, H10 and H14) were grouped with the same species as in mtDNA phylogenetic tree, *Sarotherodon nigripinnis*, *Sarotherodon knauerea* and *Sarotherodon galilaeus,* respectively (Figure 8).

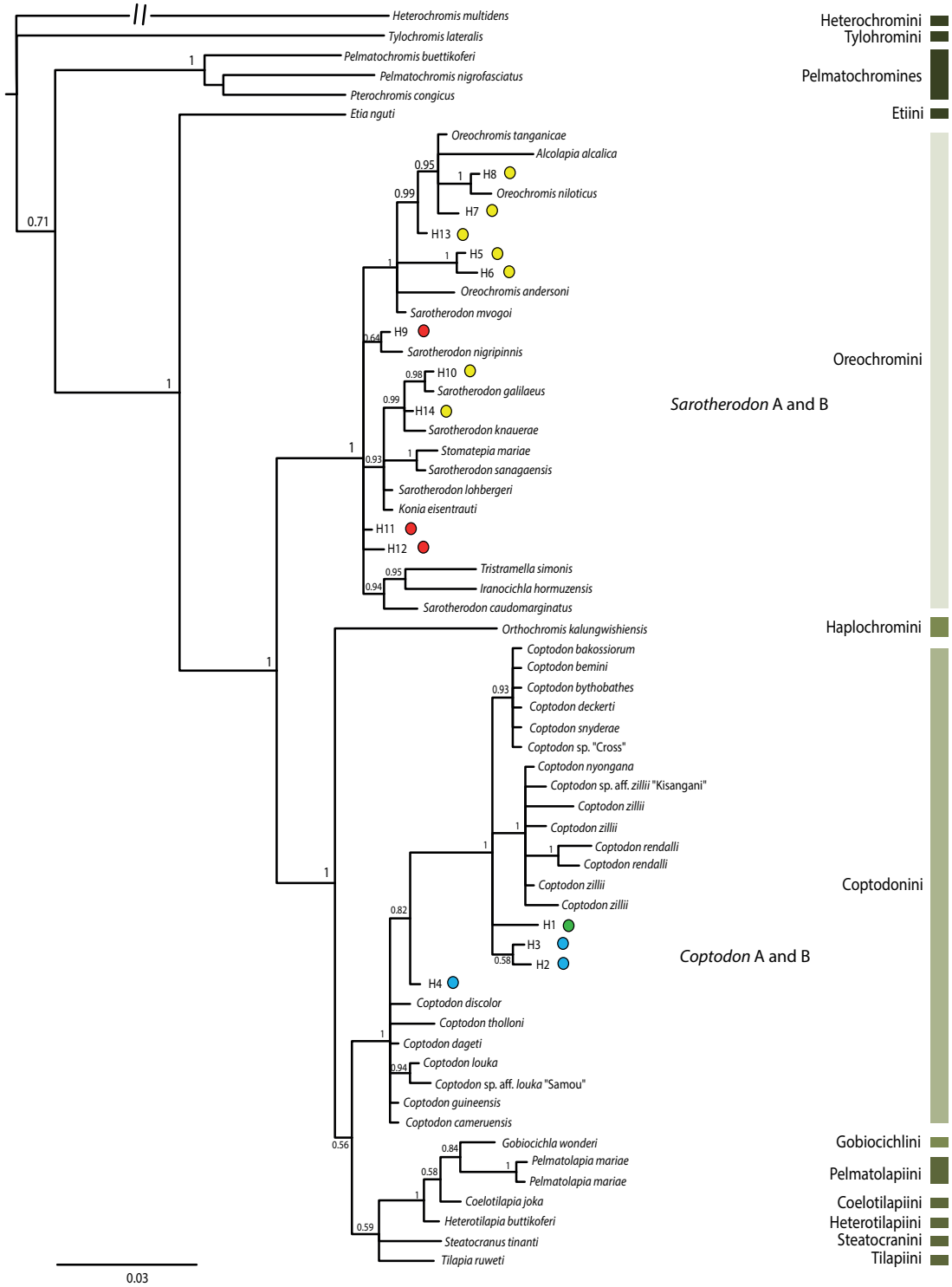


Figure 8. Phylogenetic tree representing relationships for *Coptodon* and *Sarotherodon* genera using a nDNA fragment (492bp, 1st intron of S7). A total of 14 haplotypes observed in our study and 50 taxa (Table S3, Appendix) are included. Tree is performed using Bayesian analyses in MrBayes. Each dot shows one haplotype form this study, each color represent lineage based on mtDNA.

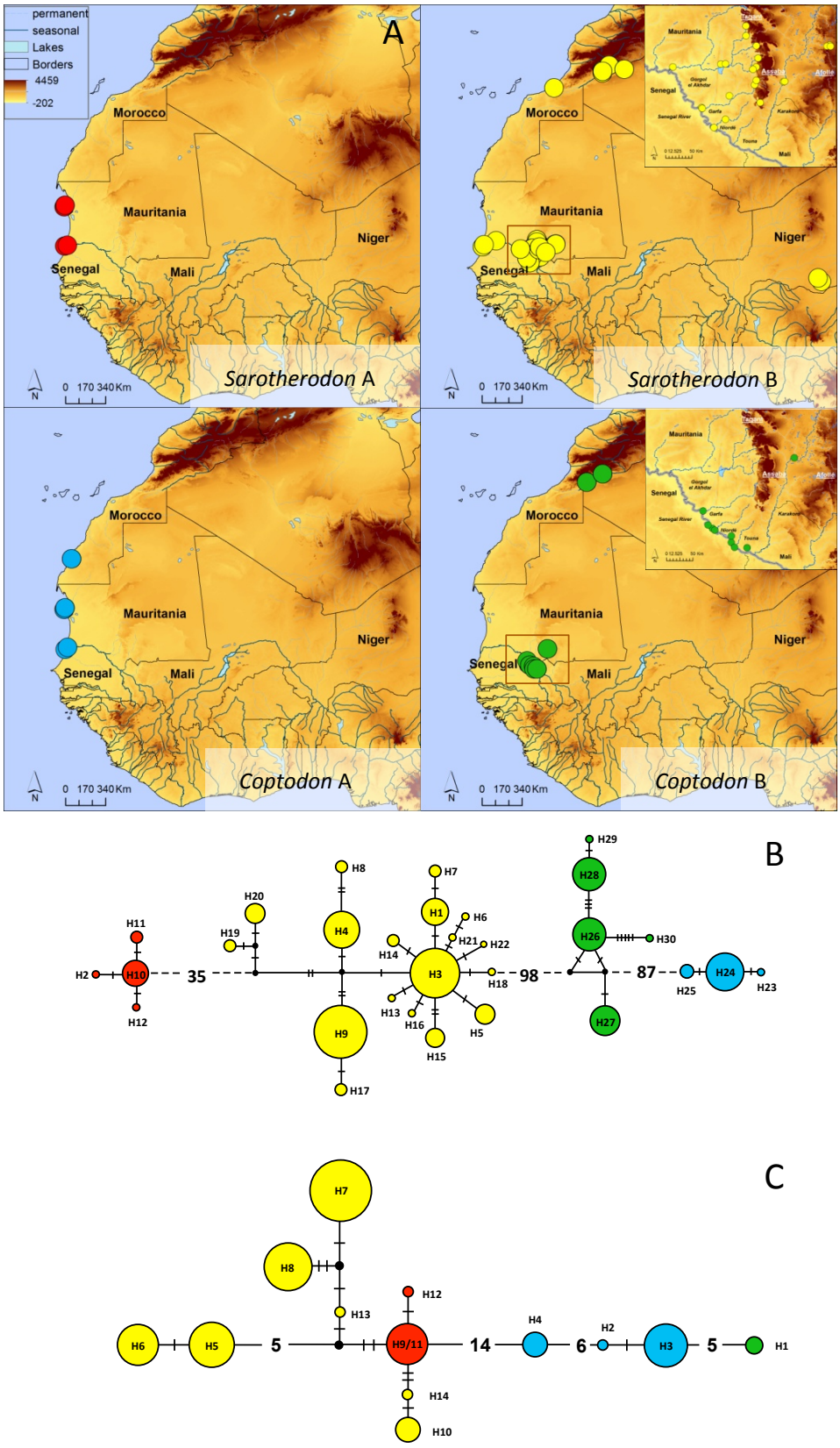


Figure 9. Haplotype networks and respective distribution maps. (A) Maps with distribution of samples from the four mtDNA lineages (*Sarotherodon* A, B and *Coptodon* A, B). Colour corresponds to the haplogroups and the dot corresponds to the geographical coordinates of sampling sites. (B) and (C) haplotype networks based on mtDNA (16S+ND2) and nDNA (1st intron of S7), respectively.

3.4 Haplotype networks of mtDNA and nDNA

As expected, the haplotype network based on mtDNA (16S+ND2) exhibited marked differences among four lineages (Figure 9B). The two putative genera, *Sarotherodon* and *Coptodon,* were separated by 96-point mutations. *Sarotherodon* B (yellow) showed the highest diversity and a part of its 19 haplotypes are distributed in a star-shaped arrangement. nDNA haplotype network, although based on a much smaller sample size, retrieve part of the differentiation exhibited by mtDNA within *Coptodon*, but *Sarotherodon* mtDNA lineages do not separate for the nDNA. Nevertheless, it is interesting that no nDNA haplotype sharing is observed among the different mtDNA lineages.

3.5 Genetic structure of *Sarotherodon* lineages

Genetic clustering analyses using STRUCTURE and all 128 *Sarotherodon* lineages samples showed the highest posterior probability ln Pr(X|K) associated with K=8 (Figure S2, **Appendix**). While following the ΔK method by Evanno *et al.* (2005), the most likely number of genetic groups (K) was two (Figure 10B). Structure results for K=2 evidenced a split between: (1) all individuals of *Sarotherodon* A with coastal distribution in Mauritania together with individuals of *Sarotherodon* B from Morocco plus eight individuals widely distributed within Mauritania inland, and (2) *Sarotherodon* B from Mauritania and Niger. The same clear pattern between two distinct genetic groups was obtained in PCoA analyses (Figure 11A), for which the first axis explains a high percentage of the total genetic variability (34.05%).

Following Evanno's method (Evanno *et al.* 2005), we tested for sub-clusters within each hierarchical genetic group (Figure 10B). The first major cluster, defined above as (1), was further subdivided into two clusters. One of the clusters was the final one, without any further subdivisions, which comprised samples from Morocco of *Sarotherodon* B (henceforth as cluster 1 and represented in light grey), the splits of the second cluster will be described next. The same division was supported by axis 2 (corresponding to 12.53% of genetic variability) in PCoA analyses (Figure 11A). The two lineages, *Sarotherodon* A and B, from coastal and inland in Mauritania, were divided into sub-clusters (Figure 10B), from the previously one, cluster 2 (henceforth represented in dark grey) and second one, which was divided

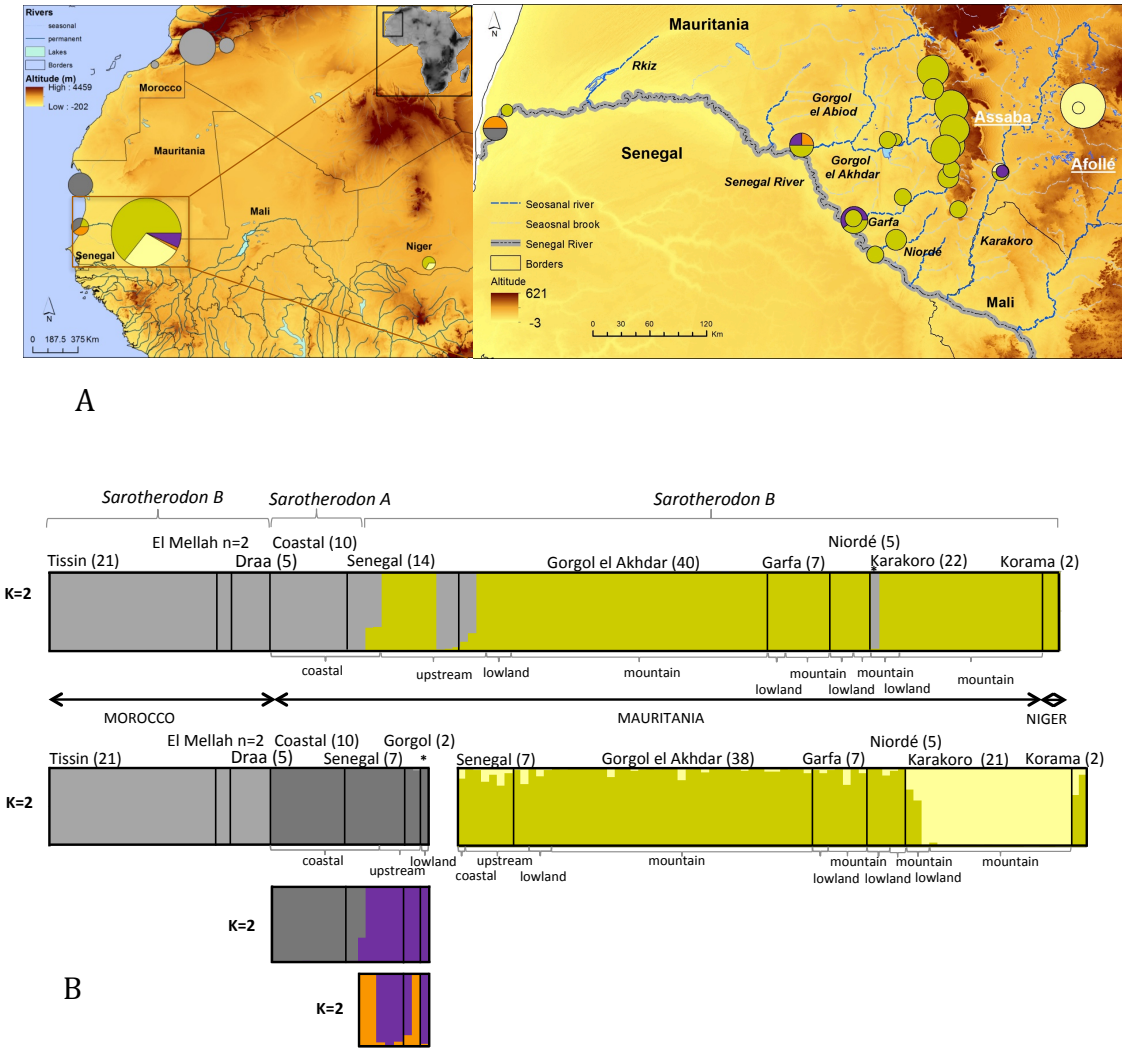


Figure 10. Sample distribution of two lineages and bar plots based on the STRUCTURE based on Evanno's method. (A) The large scale map shows West-North Africa (Morocco, Mauritania, Senegal and Niger). The smaller-scale map of Mauritania shows sub-basins, Senegal River and water bodies from Assaba and Afollé mountains. The size of pie charts is proportional to the sample frequency of each location and each color corresponds to each genetic cluster. (B) Results for the analyses using all samples of *Sarotherodon* A and B are shown in the top bar. Each inferred cluster was analyzed separately, until no structure was found. Six possible clusters were inferred. Each vertical line represents an individual and each colour represents a cluster. Individuals are grouped by basins or sub-basins, delimited by black vertical lines, and named on the top of plot bar with the number of samples shown into brackets. Mauritanian sub-basins are divided by topology shape and are shown within braces.

in the final two sub-cluster, cluster 3 (henceforth represented in purple) and 4 (henceforth represented in orange). The last two groups contained small number of individuals with wild distribution within Mauritania inland and their presence together with individuals from other genetic distinct groups (Figure 10A).

The second major cluster, defined above as (2), was sub-divided into two genetic clusters, and no further structure was found within each group (Figure 10B).

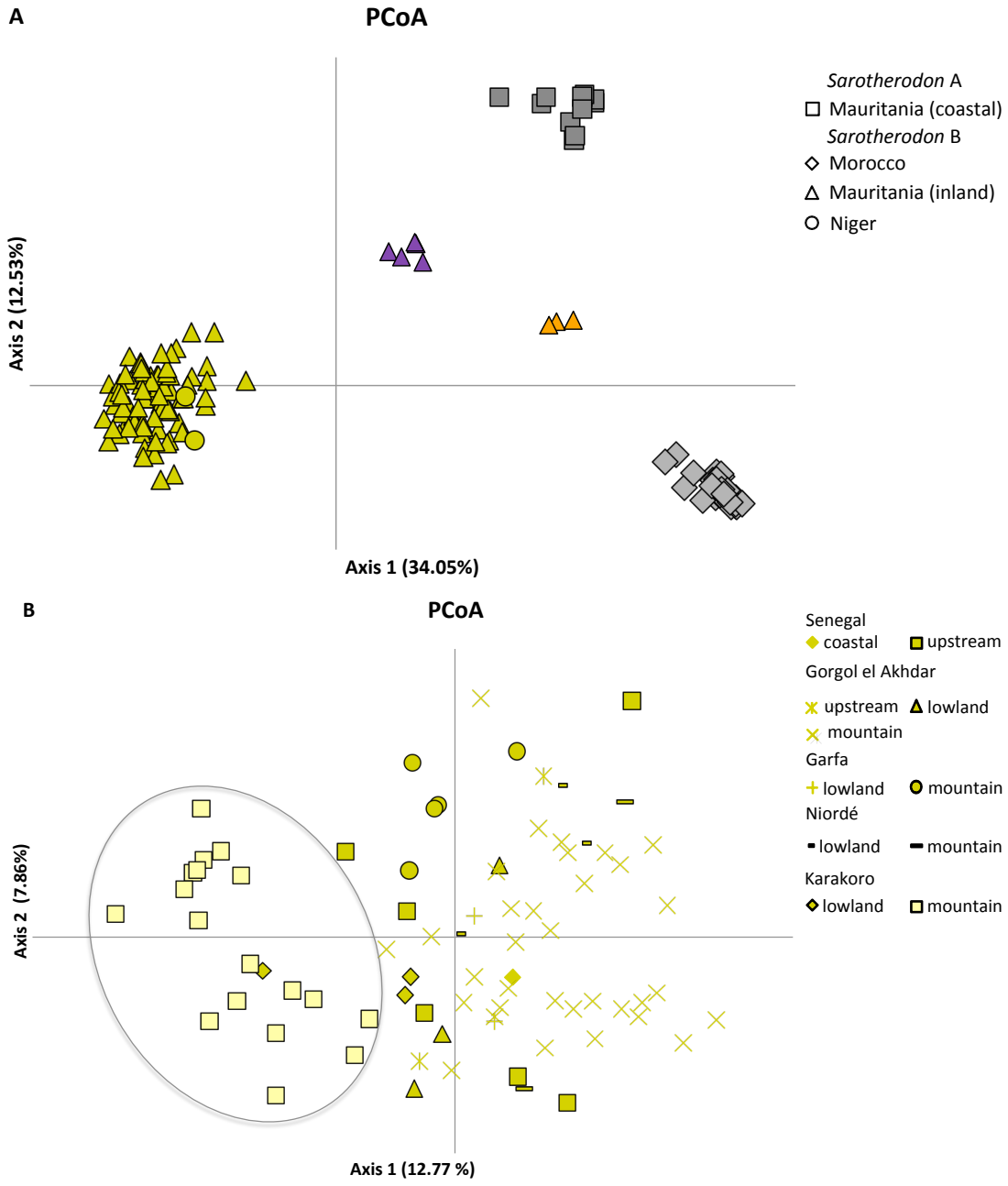


Figure 11. Principal coordinate analyses (PCoA) using the 13-microsatellite dataset for *Sarotherodon* A and B. (A) Plot of both lineages, grouping samples by countries, showing five genetically distinct groups. Each symbol represents individual and the colours correspond STRUCTURE clusters when K=5. (B) Analyses within Mauritanian cluster, showing sub-cluster of Karakoro sub-basin mountain – Afollé. Samples were grouped to sub-basins, which were divided to smaller groups due to topology. Colours represent the clusters found by STRUCTUTRE analyses**.**

The cluster 5 (henceforth represented in olive) contained the majority of individuals distributed in Mauritania inland of *Sarotherodon* B lineage and two from Niger. The cluster 6 (henceforth represented in light yellow) composed by individuals from Affolé mountain and three from Karakoro sub-basin. Two individuals had admixture of both

clusters (Figure 10A) and one more sample form another genetic group (purple). These two sub-clusters were obtained in PCoA analyses, performing it separately on the second major cluster (2) (Figure 11B), looking for the structure pattern between mountains and lowlands.

F_{ST} analyses showed the highest significant differences among obtained clusters, while all topological groups within Mauritanian cluster (cluster 5) exhibited not significant differences. Observed genetic differentiation varied between 0.1218 and 0.6413. The greatest observed genetic differentiation (F_{ST}=0.6413) was between Morocco (cluster 1) and Karakoro sub-basin (cluster 6), which are also the most geographically distant, excluding Niger. Mauritanian cluster (in olive) exhibited a genetic differentiation less pronounced to coastal *Sarotherodon* A (F_{ST}= 0.4764) than to *Sarotherodon* B from Morocco (F_{ST}= 0.5099), which can be a result of low genetic diversity within this group (Table 3). The least differentiated groups in our dataset (F_{ST}= 0.1218) are cluster 5 (olive) and 6 (light yellow; Table 2A).

F_{ST} values obtained between rivers and mountains including the combination of different sub-basins were low and not significant. The comparison for sub-basin gave higher F_{ST} values but still not significant (Table 2B).

Table 2. Pairwise F_{ST} analysis based on 12 microsatellite markers. (A) Between four clusters obtained in STRUCTURE analyses. (B) between geographical features. The genetic differences of rivers vs mountains with indication of the number of individuals compared (in brackets), and comparison between two sub-basins.

A	Cluster 2	Cluster 5	Cluster 6
Cluster 1	0.6181*	0.5099*	0.6413*
Cluster 2		0.4764*	0.6292*
Cluster 5			0.1209*

*Indicate significant values, after Bonferroni correction (P<0.005).

B **River vs mountain**	F_{ST}
Gorgol el Akhdar and Garfa (38; 7)	0.0178
Gorgol el Akhdar, Garfa and Niordé (40; 10)	0.0131
Gorgol el Akhdar vs Garfa (38; 7)	0.0415

3.6 Genetic diversity

The four genetic clusters (1, 2, 5 and 6) showed no deviation from Hardy-Weinberg equilibrium and no evidence for linkage disequilibrium in any pair of loci, after Bonferroni correction. The lowest genetic diversity was exhibited by the

Morocco cluster (light grey, Table 3) for all statistical measures analyzed (private alleles, A_R, H_O and H_E), and that was monomorphic for 7 out of 13 loci (Table S7, **Appendix**). The highest diversity values were found for clusters 5 and 4. It is interesting to observe that clusters 3 and 4 have a very low sample size (n=5 and n=3) and notwithstanding showed very high number of alleles and also private alleles, both probably evidencing an old separation of this clusters. Affolé mountain cluster (Karakoro sub-basin, light yellow) did not show the presence of any private allele.

Table 3. Genetic diversity among six clusters based on 13 microsatellite loci. The number of genotyped individuals (n), total number of alleles per locus (N_a), number of private alleles, allelic richness (A_R) based on 13 and 11 microsatellite loci, observed (H_O) and unbiased expected (H_E) heterozygosities and inbreeding coefficient (F_{IS}).
* marks clusters with low sample size for which some statistics were not calculated.

Cluster	n	N_a	Private alleles	A_R (13 loci)	A_R (11 loci)	H_O	H_E	F_{IS}
1 (light grey)	28	32	4	1.76	1.65	0.242	0.269	0.102
2 (dark grey)	12	40	15	-	2.19	0.308	0.317	0.03
3 (purple)*	5	39	14	-	-	-	-	-
4 (orange)*	3	42	10	-	-	-	-	-
5 (olive)	59	100	31	2.971	2.87	0.548	0.573	0.045
6 (light yellow)	19	41	-	2.145	1.98	0.408	0.395	-0.034

3.7 The comparison between cluster and haplotypes

Two unique genetic groups were identified based on different molecular markers. The cluster 2 (dark grey) and 3 (purple) obtained in the STRUCTURE analysis, were also identified as unique haplogroups in both haplotype networks, based on mtDNA (16S+ND2) and nDNA (1st intron of S7) (Figure 12).

All other clusters/groups are represented by shared haplotypes, having also some private haplotypes. The orange cluster grouped with purple for microsatellite data, here exhibit different relationship. Orange haplotypes were shared with Mauritanian (in olive) and Morocco (light grey) clusters (with one haplotype closest to the olive group) in mtDNA and nDNA networks, respectively. Haplotypes found at Karakoro sub-basin (light yellow) were shared with Mauritania and also with Morocco, according nDNA. These comparisons can, nevertheless, be biased by different sample size used in both haplotype networks (with much lower number of individuals for nDNA approach).

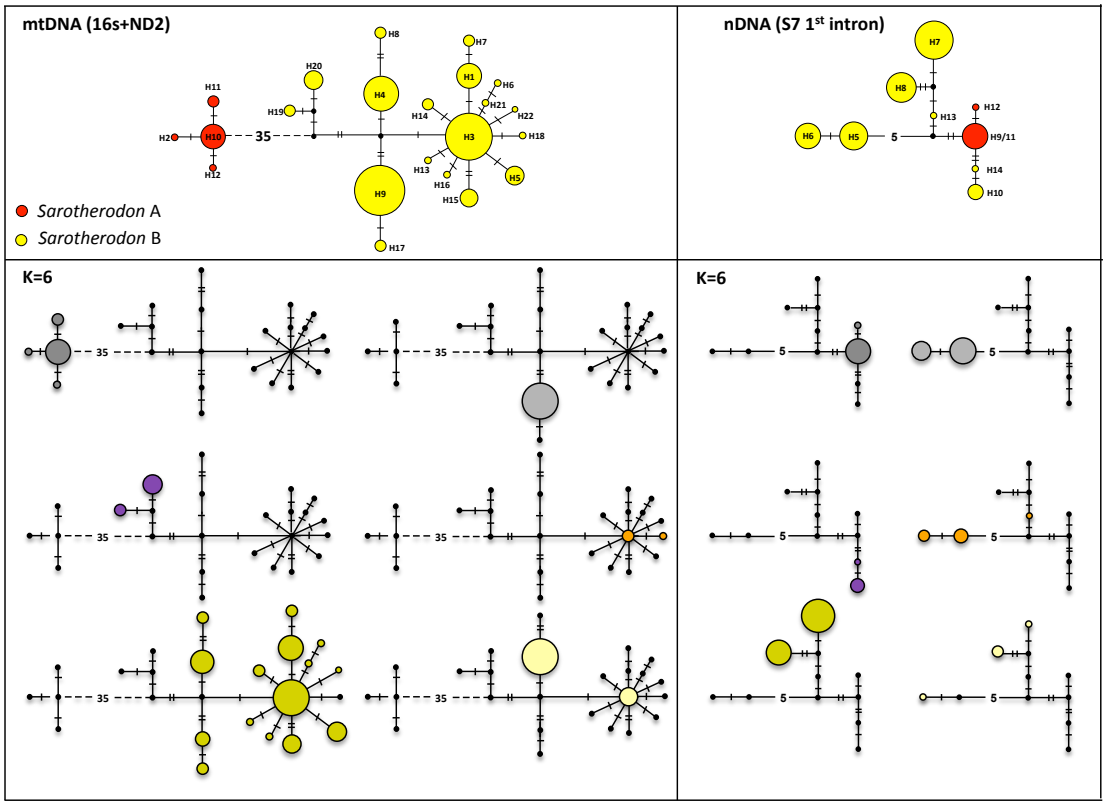


Figure 12. Haplotype networks of the samples analyzed for microsatellites. On the top, two networks based on combined mtDNA and nDNA, in the right and left, respectively. The colours represent different lineages. Below, haplotype networks are represented cluster identified by STRUCTURE analyses (K=6). The colour corresponds to each of those clusters. Circle size is proportional to sample frequency, black dots represent inferred missing haplotypes and bars/ numbers correspond to the observed mutations.

4. Discussion

4.1 Phylogenetic relationship between Tilapia lineages within African cichlids

The phylogenetic analysis based on both, mtDNA and nDNA genes, showed that the sampled Tilapias from North and West Africa are members of two tribes, Oreochromini and Coptodonini. The differentiation within each tribe was observed using mtDNA markers, as these are more sensitive to recent evolutionary events than nuclear ones (Fugure 7 and 8).

Oreochromini and Coptodonini are known to be morphologically and behaviorally divergent (Beveridge & McAndrew 2000). For example, the Oreochromini tribe species are known to exhibit biparental/parental or maternal mouthbreeder care for their offspring. Such differences may help to reinforce reproductive isolation between the two tribes. We detected parallel distribution between species from two genera. *Sarotherodon* A and *Coptodon* A with distribution in coastal areas, estuaries, while *Sarotherodon* B and *Coptodon* B are distributed in inland water bodies, with some salinity tolerances within *Sarotherodon* genera. The geographical distribution of Oreochromini and Coptodonini are overlapping, but not identical, suggesting that allopatric processes may have been important in their diversification.

Some lineages were identified at species level, combining the phylogenetic and published/ unpublished information, while others lineages were unresolved, as complex species. The lineage *Sarotherodon* A the most probably is *Sarotherodon melanotheron (*Rüppell, 1852) species, which is sister species to *Sarotherodon nigripinis* (Falk et al. 2003), as were obtained in phylogenetic analyses based on both, mtDNA and nDNA, molecular approaches (Figure 7 and 8). *S. melanotheron* and *S. nigripinis* are coastal lineages with distribution from Senegal to Benin and from the mouth of the Congo to the Lower Kouilou, respectively. With probable divergence of lineages in early Pleistocene (Falk *et al.* 2003). The black-chinned tilapias*, S. melanotheron,* are known as marine species tolerant to wide range of salinity (Lemarié *et al.* 2004). Moreover, Kide *et al.* (unpublished) described Tilapias from PNBA Mauritania (common samples with our study), as *S. melanotheron* based on morphometric parameters, meristic counts and phylogenetic analyses, using four genetic markers.

While another lineage *Sarotherodon B* is more complex, where two haplotypes, in both phylogenetic analyses, are grouped with *Sarotherodon knauerae* and *Sarotherodon galilaeus.* The rest of the haplotypes exhibited monophyletic sub-clade according to mtDNA data, while due to nDNA phylogenetic data haplotypes were grouped with a several different species. There are records of the subspecies *Sarotherodon galilaeus galilaeus* distribution in Adrar (Mauritania) (Figure 5) (Lévêque 1990) and in the lowlands of Morocco (Le Berre 1989), close to our study area. Moreover, according to the nDNA data, the haplotype H10 grouped with *Oreochromis niloticus* with a high support value (Figure 8). According to species distribution collected by Trewavas (1983) the *Oreochromis niloticus was* recognized in Senegal River. Besides the species was introduced in the expansion of Senegal River (M.M.J. Vincke, per. Comm. to Pullin (1988)).

Coptodon A, which was a coastal lineage based on mtDNA analyses, has been described as the *Coptodon guineensis* species complex in Kide *et al.* (unpublished). *Coptodon guineensis* appears to exist in sympatry with other genera, including *Sarotherodon melanotheron* (in our study named as *Sarotherodon A*), in Mauritanian coastal areas. Understanding the genetic relationships among A and B lineages in *Sarotherodon* genera, on which are focus our study, may require determining the relationships between parallels lineages such as lineages A and B (in our study) in *Coptodon* genera.

Individuals that were found to be part of the inland mtDNA lineage, *Coptodon* B, did not form a monophyletic clade in our phylogenetic analyses based on nDNA data. The lack of concordance between the results of mtDNA and nDNA phylogenetic analyses may have been due to using too few unlinked nuclear loci in the latter. However, one haplotype (H27, Figure 7) are grouped with *Coptodon zillii*. Furthermore, there are some records about *Tilapia* (*Coptodon) zillii* distribution in locations close to our study area (up to Atlas mountains and West Africa) (Le Berre 1989; Lévêque 1990). In addition, it was recently mentioned about that redbelly tilapia, *Coptodon zillii*, occurrs in the Draa basin (Morocco) together with *Oreochromis aureus* commonly named as blue tilapia (Clavero *et al.* 2014). These reports conflict with the results of our mtDNA phylogenetic analyses, which do not indicate *Sarotherodon* B as *Oreochromis* (found in the same locations), while based on one nuclear locus one of the haplotypes (H8 in Figure 8) is grouped with *Oreochromis niloticus*. However, these conclusions based on two mtDNA and one nDNA loci requires further testing. We therefore tentatively conclude that both genera are closely related and are from the same tribe. It is important to highlight,

that recent phylogenetic studies of African cichlids were based on multi-gene data and most of the samples are from the West, center and South of Africa. Therefore, it is possible that some new species within the identified tribes are in our samples that have not yet been described or are not included in the previous phylogenetic analyses.

Our main goal of the phylogenetic analyses was to detect the species occurrence in Mauritania water bodies and to perform further population genetic analyses. We detected two lineages of different genera, *Sarotherodon* B and *Coptodon* B, distributed in these areas. While the majority of species presented in lowland and mountains was from *Sarotherodon* B lineage (Figure 9A). *Coptodon* B were present in lowland of Karakoro sub-basin, but majority of samples were from Senegal River and a few from downstream of the sub-basins. The lineages *Sarotherodon* A and *Coptodon* B are present in sympatry in some locations, according to our sample distribution map (Figure 9). But any fishes from *Coptodon* B were found in mountain water bodies. Perhaps they might occur together in more areas, our results can be biased due to lack of sampling, which is correlated with climate conditions. These lineages belong to different genera and tribe, so we can reject the possibility of interaction between them.

In any case our results showed that the broadest distribution in Mauritanian lowlands and mountains is represented by fishes from *Sarotherodon* B lineage. The further population studies were performed on this lineage and for this reason the novel microsatellites were developed. According to the close genetic relationship between *Sarotherodon* A and B, based on nDNA sequences (Figure 9C), the second lineage also was included to the further analyses.

4.2 Inferring historical events

The detection of the recombination can give the idea about the gene flow event before the species diverged, to adapt to environmental conditions. The recombination was detected in samples from two lineages, *Sarotherodon* B and *Coptodon* B (Figure S1, **Appendix**). Samples were collected in Mauritania. Likely, Mauritania acted as secondary contact zone to those individuals; it could be related with the beginning of dry period, which pushed individuals in restricted patches.

The haplotype network based on mtDNA represented the same four lineages as in phylogenetic analyses. The two genera *Sarotherodon* and *Coptodon* were distinct by highest mutation number 98 (Figure 9). The star-shaped haplotype

network of *Sarotherodon* B (yellow) lineage distributed in Mauritania, suggested it may have experienced a recent range expansion. This range expansion may have been the result of climate and habitat changes and that opened up the waterways and allowed for dispersal to occur. It is possible that similar patterns exist in *Coptodon* B distributed in Mauritania body water, but our limited sample size did not allow us to test this hypothesis.

Some of the relationships among the individuals we studied were found to be different based on the mtDNA and nDNA sequence phylogenetic analyses (Figure 9 B, C). *Sarotherodon* A exhibited the intermediate relation within *Sarotherodon* B haplotypes based on nDNA data, separating H14 and H10 from the rest (Figure 9C). In the phylogenetic analyses *Sarotherodon* A (coastal) was grouped with *Sarotherodon nigripinis* (Figure 8), which diverged about 10 Mya (see Figure 3 in Schwarzer *et al.* (2009)). The sister lineage diverged into two species (*Sarotherodon sp. aff. galilaeus (*recently corrected to *Sarotherodon knauerae)* and *Stomatepatia mariae*) about 2.5 Mya. One group of the inland lineage *Sarotherodon* B (H14 in Figure 8) is grouped with *Sarotherodon knauerae*. Because of the lower mutation and slower fixation rates of nuclear loci compared to mtDNA, they can sometime provide information about older evolutionary events. We therefore hypothesize that some inland groups (from *Sarotherodon* B lineage) diverged from coastal lineages.

4.3 Population structure and genetic diversity in West Africa

The genetic differences may be due to allopatric divergence or selection among biogeographic regions. The final number (K=6) of the genetic groups obtained within *Sarotherodon* A and B, across the North-West Africa using 13 microsatellite loci were six, supported by structure analyses based on Evanno's method and PCoA, as well inferring their differences by F_{ST} statistics. Two clusters, Morocco (light grey) and coastal of Mauritania (dark grey), of different lineages (based on mtDNA), showed high (F_{ST}) significant genetic differences. The same two groups highly differ from other two genetic groups, Mauritania (olive) and Karakoro sub-basin (light yellow). The barrier between coastal and fresh waters can be the concentration of salinity. However, F_{ST} is sensitive to low heterozigosity. The high F_{ST} values obtained on Morocco population, was also due to monomorphic loci. The coastal group of *Sarotherodon* A showed missing data for one locus. Perhaps due to some mismatch in the flanking sequences, because the primers were designed for

the close related (Wang *et al.* 2012) lineage, *Sarotherodon* B which is from the same tribe.

The Eastern Mauritanian mountain population of *Sarotherodon* B appears to be the most recent isolated. The population of *Sarotherodon* B fishes from Affolé mountain (in Karakoro sub-basin) exhibited population genetic structure isolated from Western Mauritania. Two genetic clusters obtained in Mauritania, olive and light yellow, had a high F_{ST} value (0.1209), common at population level (Cornuet *et al.* 1999). The same population isolation results was obtained by Velo-Antón *et al.* (2014) in *Crocodylus suchus* population from Affolé mountain. This area has minimal water availability comparing with the west-ward (Campos *et al.* 2012). However, some individuals were found in Karakoro sub-basin, with admixture of both populations. Unfortunately, we couldn't get any samples from Karakoro sub-basin in downstream from Afollé mountain, to get better idea about population dynamics. Besides, any private alleles were observed in this population. This may have been the result of a recent bottleneck, which can decrease the number of alleles and heterozygosity in a population, and increase the rate at which rare alleles are lost, especially in small isolated populations (Maruyama & Fuerst 1985; Nei *et al.* 1975). We did not test for population bottlenecks in this case due to small sample size and the lack of polymorphism in the population. However, our results from analyses of three types of genetic data (microsatellites, nDNA and mtDNA), suggest that it is an isolated population.

No evidence for genetic sub-structure within the *Sarotherodon* B from Mauritania (olive cluster) was found. Genetic differentiation analyses were performed on many different combinations of groups, to detect gene flow or weak structure and their relation with hydraulic networks. The coastal group exhibited very distinct genetic relation to lowland and mountain putative populations, according to F_{ST} analyses (Table 2A). Either, coastal population is considered as other lineage, *Sarotherodon* A based on mtDNA data. While F_{ST} analyses performed within Mauritanian cluster (olive) on rivers versus mountains, including one up to three sub-basins and between two sub-basins, any significant differences were observed. The highest value of F_{ST} was obtained between Gorgol el Akhdar vs Garfa sub-basins. Moreover, the metapopulations of *C. suchus* crocodiles were isolated between sub-basins of Mauritania, with hydraulic connections within. On the contrary, the fish samples of two clusters, olive and light yellow, were found in Karakoro sub-basin. Topology is specifically important by constraining their ability to disperse (Fagan 2002) and also is very important to know about traits of species (Leibold *et al.* 2004). While these two taxa, fishes and crocodiles have different dispersal abilities,

reproduction and lifespan. Our obtained data can be better interpreted by increasing the number of individuals, to obtain statistically significant values and to cover the sampling gaps, especially from downstream of Karakoro, to have better sense about population dynamics.

Moreover, two genetic distinct groups (purple and orange; Figure 10 and 11) were identified within Mauritania and each was found to be genetically isolated from the rest of the Mauritanian populations with which they are sympatric. The genetic differentiation wasn't obtained among these groups, due to small sample size (5 and 3). However, within these two groups were obtained high genetic diversity values. The orange population exhibited the highest value of allelic richness and purple as well, after Mauritanian cluster (in olive). The 3 individuals of orange cluster are distributed in coastal area of Senegal River and in the downstream of Gorgol sub-basin. While individuals from purple cluster are distributed in mountains, in Karakoro sub-basin and in downstream of Gorgol and Garfa sub-basins. These two clusters were found in sympatry, as well as with individuals from other clusters (Figure 10). The individuals of the purple cluster, as a different group, were also obtained based on other molecular markers. While orange group was identified, based only on microsatellite data. Perhaps, these groups should be separated by assortative mating, due to divergence in feeding apparatus, or colour preference common in cichlids from Lake Malawi (Kocher 2004). It is important to increase the sample size, because it seems that these individuals was caught by chance. The morphology data analyses on these specimens could be helpful.

4.4 Historical and contemporary structure of Tilapias

We can see some pattern of sympatry in Mauritania, which have been involved at time. The purple cluster defined based on microsatellites data, distinct from the Mauritania (olive and light yellow), showed similar pattern based on mtDNA and nDNA sequences. The factors acting as barriers, should to be involved long time ago, according the resolution of markers. Besides, close relation was detected between purple and dark grey (coastal) based on the both nDNA marks, 1st intron of S7 and microsatellite loci, without F_{ST} support. Perhaps it is possible some interaction between coastal and fresh waters fishes from the same genera, knowing that *Sarotherodon melanotheron* exhibit tolerance to a wide range of salinity (Lemarié *et al.* 2004). Moreover, the salinity of water can decrease by increased

water fluidity during rainy seasons (Panfili *et al.* 2004; Panfili *et al.* 2006). However, the mating traits, as body size, colour, behavior, is high variable in cichlid species (Kocher 2004), can drive to early divergence in closely related species**.**

There is probability of subspecies/ close species relationship within *Sarotherodon* B lineage**.** The previous mentioned purple group with two haplotypes H19 and H20 (phylogenetic tree of the mtDNA) are grouped with *Sarotherodon sp*., in the paraphyletic relation to the rest haplotypes of *Sarotherodon* B lineage. Two haplotypes H10 and H14 (phylogenetic tree of the nDNA) are grouped with the same species. Furthermore, both haplotypes networks showed that purple group exhibited an unique haplogroup and no shared haplotypes. This could be evidence of closely related species/ subspecies.

While orange cluster perhaps exhibit more recent stage of sympatry, comparing with purple cluster. The orange cluster was grouped with purple and dark grey (coastal) clusters, based on STRUCTURE and PCoA analyses. While based on other markers, shared haplotypes with Mauritania or Morocco. Any different pattern was detected in phylogenetic analyses. This differentiation involves markers with sensibility to more recent events – microsatellites. It might be more recent divergent events, comparing with purple cluster. However, the orange group is more distinct than light yellow due to high number of private allele. Since individuals of orange cluster occur in the same areas, one of the main driver of divergence might be assortative mating (Stelkens & Seehausen 2009).

The rest clusters of *Sarotherodon* B (light grey, olive and light yellow) have shared alleles in both haplotype networks. And in microsatellites they are divided in three clusters, with the most distinct Morocco group and differentiation between groups of Mauritania (olive) and Karakoro sub-basin (light yellow). The haplotype network didn't show any pattern of hydraulic network biogeography (data not shown), the same as based on microsatellite data.

The most of the hypothesis are not supported due to small sample size, which is a barrier to perform analyses, by increasing the bias to interpretation of data. Also is important to get samples from more locations, and more markers, get better idea about phylogenetic data and exhibit the fishes population dynamics populations within hydraulic network. More markers, as nuclear gene and high polymorphic microsatellite loci, could to increase the power of data.

5. Conclusions

Mitochondrial and nuclear data in Tilapia fishes, together with sequences of other African cichlids, allowed us to study them at phylogenetic level. Inferring the phylogenetic relation between Tilapias in West and North of Africa, especially within Mauritania inland, and comparing with other cichlids. One of our findings was a strong phylogenetic structure in West-North Africa Tilapias of two groups, corresponding to two tribes, Oreochromini and Coptodonini, which were divided into two further groups to give a total of four mitochondrial lineages, *Sarotherodon* A and B, and *Coptodon* A and B. Each genera contains the lineage with coastal (A) and inland (B) distribution, with some salinity tolerance within *Sarotherodon* genera. *Sarotherodon* B was over distributed in Mauritania mountains and lowlands, while *Coptodon* B mainly appear in Senegal River or in some downstream of sub-basins. Nuclear data seem to indicate a closer relation within and among lineages, and some intermediate relation of *Sarotherodon* A within *Sarotherodon* B lineage, was detected from haplotype network. Some haplotypes from two lineages were hypothesized to belong to known species according to phylogenetic and published data, where *Sarotherodon* A is close to *Sarotherodon melanotheron* and *Coptodon* B is close to *Coptodon zillii*. While another two lineages are more complex. Also, the recombination events were detected in samples collected from Mauritania, encompassing *Sarotherodon* B and *Coptodon* B haplotypes, which might be the result of secondary contact zone. More data will be necessary to confirm the relation among lineages at phylogenetic level and recombination events.

Thirteen microsatellite loci data, allowed us to identify two genetic groups in Mauritania fresh waters. Afollé mountain exhibited isolated fish population of *Sarotherodon* B (with a minimal water availability) from the Western Mauritania, with some admixture of both clusters in Karakoro sub-basin. None of the significant statistical support was obtained by examining the differentiation between hydraulic networks in Mauritania, at smaller geographical scale. It seems that Tilapias population dynamics could be related with hydraulic network within each sub-basin, except Karakoro sub-basin. More sample will be needed to improve the study, and especially from some locations. The inclusion of landscape genetics studies and the use of species behavior knowledge may help as well.

More four lineages were identified of *Sarotherodon* A and B. Two clusters represented biogeographic differentiation to Mauritania inland populations: (1) North

Africa, Morocco of Sarotherodon B and (2) Coastal, Mauritania Sarotherodon A. While the most surprising results were two clusters of *Sarotherodon* B, with very low sample size, widely distributed within Mauritania inland with some cases of sympatry. Both clusters were distinct from the rest of Mauritanian Tilapias, according to different genetic data, which could be the result of different stages in divergences. In this case is important samples size; since it was used a minimum number and morphological studies could give a better sense.

6. References

Adams J, Faure H (2004) Review and Atlas of Palaeovegetation - Preliminary land ecosystem maps of the world since the Last Glacial Maximum. Quaternary Environment Network

Allendorf FW, Luikart GH (2007) *Conservation and the genetics of populations* Blackwell Publishing Ltd.

Avise JC (2000) *Phylogeography: the history and formation of species* Harvard University Press.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution* **16**, 37-48.

Beveridge MC, McAndrew BJ (2000) *Tilapias: biology and exploitation* Springer.

Blacket M, Robin C, Good R, Lee S, Miller A (2012) Universal primers for fluorescent labelling of PCR fragments–an efficient and cost‐effective approach to genotyping by fluorescence. *Molecular ecology resources* **12**, 456-463.

Blouin MS, Phillipsen IC, Monsen KJ (2010) Population structure and conservation genetics of the Oregon spotted frog, Rana pretiosa. *Conservation Genetics* **11**, 2179-2194.

Bohonak AJ (1999) Dispersal, gene flow, and population structure. *Quarterly review of biology*, 21-45.

Brito JC, Godinho R, Martínez‐Freiría F*, et al.* (2014) Unravelling biodiversity, evolution and threats to conservation in the Sahara‐Sahel. *Biological Reviews*.

Brito JC, Martínez-Freiría F, Sierra P, Sillero N, Tarroso P (2011) Crocodiles in the Sahara desert: an update of distribution, habitats and population status for conservation planning in Mauritania. *PLoS ONE* **6**, e14734.

Brown J, Lomolino MV (1998) Biogeography (2nd edn). *Sunderland MA: Sinauer Associates*.

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the national Academy of Sciences* **76**, 1967-1971.

Campos JC, Sillero N, Brito JC (2012) Normalized difference water indexes have dissimilar performances in detecting seasonal and permanent water in the Sahara-Sahel transition zone. *Journal of Hydrology*.

Cañón J, Alexandrino P, Bessa I*, et al.* (2001) Genetic diversity measures of local European beef cattle breeds for conservation purposes. *Genetics Selection Evolution* **33**, 311-332.

Chakrabarty P (2004) Cichlid biogeography: comment and review. *Fish and fisheries* **5**, 97-119.

Chow S, Hazama K (1998) Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology* **7**, 1255-1256.

Clavero M, Esquivias J, Qninba A*, et al.* (2014) Fish invading deserts: non‐native species in arid Moroccan rivers. *Aquatic Conservation: Marine and Freshwater Ecosystems*.

Collin R, de Maintenon M (2002) Integrative approaches to biogeography: Patterns and processes on land and in the sea. *Integrative and comparative biology* **42**, 911-912.

Cooper A, Shine T, McCann T, Tidane D (2006) An ecological basis for sustainable land use of Eastern Mauritanian wetlands. *Journal of Arid Environments* **67**, 116-141.

Cornuet J-M, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989-2000.

Dekeyser PL, Villiers A (1956) Contribution à l'étude du peuplement de la Mauritanie. Notations écologiques et biogéographiques sur la faune de l'Adrar. *émoires de l'Institut Français d'Afrique Noire* **44**, 9-222.

DeLong DC (1996) Defining biodiversity. *Wildlife Society Bulletin*, 738-749.

Dickman CR, Pimm SL, Cardillo M (2007) The pathology of biodiversity loss: the practice of conservation. *Key topics in conservation biology*, 1-16.

Drummond A, Ashton B, Cheung M*, et al.* (2009) Geneious v4.8.5, p. Available from. http://www.geneious.com.

Dunz AR, Schliewen UK (2013) Molecular phylogeny and revised classification of the haplotilapiine cichlid fishes formerly referred to as “Tilapia”. *Molecular phylogenetics and evolution* **68.1**, 64-80.

Durant S, Wacher T, Bashir S*, et al.* (2014) Fiddling in biodiversity hotspots while deserts burn? Collapse of the Sahara's megafauna. *Diversity and Distributions* **20**, 114-122.

Dynesius M, Jansson R (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the national Academy of Sciences* **97**, 9115-9120.

Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359-361.

Edwards S, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**, 1839-1854.

Etienne RS, Olff H (2004) A novel genealogical approach to neutral biodiversity theory. *Ecology Letters* **7**, 170-175.

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.

Ezcurra E, Warning UNEPDoE, Assessment (2006) *Global Deserts Outlook* United Nations Environment Programme.

Fagan WF (2002) Connectivity, fragmentation, and extinction risk in dendritic metapopulations *Ecology* **83**, 3243-3249.

Falk TM, Teugels GG, Abban EK, 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.

Franck J, Wright JM, McAndrew B (1992) Genetic variability in a family of satellite DNAs from tilapia (Pisces: Cichlidae). *Genome* **35**, 719-725.

Franck JP, Kornfield I, Wright JM (1994) The utility of SATA satellite DNA sequences for inferring phylogenetic relationships among the three major genera of tilapiine cichlid fishes. *Mol Phylogenet Evol* **3**, 10-16.

Frankham R (2005) Genetics and extinction. *Biological Conservation* **126**, 131-140.

Gaston KJ (2000) Global patterns in biodiversity. *Nature* **405**, 220-227.

Gonçalves DV, Brito JC, Crochet P-A*, et al.* (2012) Phylogeny of North African Agama lizards (Reptilia: Agamidae) and the role of the Sahara desert in vertebrate speciation. *Molecular phylogenetics and evolution* **64**, 582-591.

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* **4**, 379-391.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9. 3).

Haas E, Bartholomé E, Combal B (2009) Time series analysis of optical remote sensing data for the mapping of temporary surface water bodies in sub-Saharan western Africa. *Journal of Hydrology* **370**, 52-63.

Hillis DM, Moritz C, Mable BK (1996) *Molecular Systematics* Second Edition, Sinauer Associates, Suderland.

Holmes JA (2008) How the Sahara became dry. *Science* **320**, 752-753.

Julian D. Olden, Mark J. Kennard, Fabien Leprieur*, et al.* (2010) Conservation biogeography of freshwater fishes: recent progress and future challenges. *Diversity and Distributions* **16**, 496-513.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* **7**, 1225-1241.

Kide NG, Dunz A, Agnèse J-F*, et al.* (unpublished) A new cichlid species in West African marine waters. Coptodon guineensis, a species complex?

Klett V, Meyer A (2002) What, if anything, is a Tilapia?—mitochondrial ND2 phylogeny of tilapiines and the evolution of parental care systems in the African cichlid fishes. *Molecular biology and evolution* **19**, 865-883.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics* **5**, 288-298.

Kornfield I, Ritte U, Richler C, Wahrman J (1979) Biochemical and cytological differentiation among cichlid fishes of the Sea of Galilee. *Evolution*, 1-14.

Kröpelin S, Verschuren D, Lézine A-M*, et al.* (2008) Climate-driven ecosystem succession in the Sahara: the past 6000 years. *Science* **320**, 765-768.

Lacaux J, Tourre Y, Vignolles C, Ndione J, Lafaye M (2007) Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote Sensing of Environment* **106**, 66-74.

Lanfear R, Calcott B, Ho SY, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution* **29**, 1695-1701.

Le Berre M (1989) Faune du Sahara. 1. Poissons - Amphibiens - Reptiles, Raymond Chabaud-Lechevalier, Paris., p. 332.

Le Houérou HN (1997) Climate, flora and fauna changes in the Sahara over the past 500 million years. *Journal of Arid Environments* **37**, 619-647.

Leibold MA, Holyoak M, Mouquet N*, et al.* (2004) The metacommunity concept: a framework for multi‐scale community ecology. *Ecology Letters* **7**, 601-613.

Lemarié G, Baroiller JF, Clota F, Lazard J, Dosdat A (2004) A simple test to estimate the salinity resistance of fish with specific application to O. niloticus and S. melanotheron. *Aquaculture* **240**, 575-587.

Lévêque C (1990) Relict tropical fish fauna in Central Sahara. *Ichthyol. Explor. Freshwaters* **1**, 39-48.

Lovejoy TE (1980) Discontinuous wilderness: minimum areas for conservation. *Parks* **5**, 13-15.

Malausa T, Gilles A, Meglecz E*, et al.* (2011) High‐throughput microsatellite isolation through 454 GS‐FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular ecology resources* **11**, 638-644.

Marsjan P, Oldenbroek J (2007) Molecular Markers, a tool for exploring gene diversity. *The state of the world's animal genetic resources for food and agriculture. FAO research report. FAO, Rome*, 359-379.

Maruyama T, Fuerst PA (1985) Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* **111**, 675-689.

McAndrew B, Majumdar K (1984) Evolutionary relationships within three Tilapiine genera (Pisces: Cichlidae). *Zoological Journal of the Linnean Society* **80**, 421-435.

Meglécz E, Costedoat C, Dubut V*, et al.* (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* **26**, 403-404.

Mitchell M, Gonder M (2013) Primate speciation: A case study of African apes. *Nature Education Knowledge* **4**, 1.

Monod T (1951) Contribution à l'étude du peuplement de la Mauritanie. Poissons d'eau douce. . *Bulletin de l'Institut Fondamental d'Afrique Noire* **13**, 802-812.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, 1-10.

Nicholson SE (1995) Sahel, West Africa. *Encyclopedia of Environmental Biology* **3**, 261-275.

Panfili J, Mbow A, Durand J-D*, et al.* (2004) Influence of salinity on the life-history traits of the West African black-chinned tilapia (Sarotherodon melanotheron): comparison between the Gambia and Saloum estuaries. *Aquatic Living Resources* **17**, 65-74.

Panfili J, Thior D, Ecoutin JM, Ndiaye P, Albaret JJ (2006) Influence of salinity on the size at maturity for fish species reproducing in contrasting West African estuaries. *Journal of Fish Biology* **69**, 95-113.

Park S (2001) The Excel Microsatellite Toolkit (version 3.1). Animal Genomics Laboratory, UCD, Ireland.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Phillipsen IC, Lytle DA (2013) Aquatic insects in a sea of desert: population genetic structure is shaped by limited dispersal in a naturally fragmented landscape. *Ecography* **36**, 731-743.

Pimm SL (2008) Biodiversity: Climate Change or Habitat Loss — Which Will Kill More Species? *Current Biology* **18**, R117-R119.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Pullin RS (1988) Tilapia genetic resources for aquaculture. *The WorldFish Center Working Papers*.

Rambaut A, Drummond A (2007) Tracer v1. 5. htt p. *beast. bio. ed. ac. uk/Tracer*.

Ronquist F, Teslenko M, van der Mark P*, et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539-542.

Rozas J, Rozas R (1995) DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. *Computer applications in the biosciences: CABIOS* **11**, 621-625.

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**, 17.

Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. *Genetics* **156**, 879-891.

Schwarzer J, Misof B, Tautz D, Schliewen UK (2009) The root of the East African cichlid radiations. *BMC Evolutionary Biology* **9**, 186.

Simberloff D, Farr JA, Cox J, Mehlman DW (1992) Movement corridors: conservation bargains or poor investments? *Conservation Biology* **6**, 493-504.

Sodsuk P, McAndrew B (1991) Molecular systematics of three tilapüne genera Tilapia, Sarotherodon and Oreochromis using allozyme data. *Journal of Fish Biology* **39**, 301-308.

Stelkens R, Seehausen O (2009) Phenotypic divergence but not genetic distance predicts assortative mating among species of a cichlid fish radiation. *Journal of evolutionary biology* **22**, 1679-1694.

Sültmann H, Mayer W (1997) Reconstruction of cichlid fish phylogeny using nuclear DNA markers.

Sultmann H, Mayer WE, Figueroa F, Tichy H, Klein J (1995) Phylogenetic analysis of cichlid fishes using nuclear DNA markers. *Mol Biol Evol* **12**, 1033-1047.

Swingland IR (2000) Biodiversity, definition of. *Encyclopedia of biodiversity* **1**, 377-391.

Tokeshi M (2009) *Species coexistence: ecological and evolutionary perspectives* John Wiley & Sons.

Toupet C (1966) Etude du milieu physique du massif de l'Assaba (Mauritanie): introduction à la mise en valeur d'une région sahélienne.

Trape S (2009) Impact of Climate Change on the Relict Tropical Fish Fauna of Central Sahara: Threat for the Survival of Adrar Mountains Fishes, Mauritania. *PLoS ONE* **4**, e4400.

Trape S (2013) A study of the relict fish fauna of northern Chad, with the first records of a polypterid and a poeciliid in the Sahara desert. *Comptes Rendus Biologies*.

Trewavas E (1983) Tilapiine fishes of the genera Sarotherodon, Oreochromis and Danakilia.

Vale CG (2010) Biogeography and Conservation of African Sahel Biodiversity. *MSc thesis in Biodiversity, Genetics and Evolution. Faculty of Sciences. University of Porto, Portugal*, pp. 102.

Vale CG, Álvares F, Brito JC (2012a) Distribution, suitable areas and conservation status of the Felou gundi (Felovia vae Lataste 1886). *mammalia* **76**, 201-207.

Vale CG, Tarroso P, Campos JC (2012b) *Distribution, suitable areas and conservation status of the Boulenger's agama (Agama boulengeri, Lataste 1886)* Amphibia-Reptilia.

Velo-Antón G, Godinho R, Campos JC, Brito JC (2014) Should I Stay or Should I Go? Dispersal and Population Structure in Small, Isolated Desert Populations of West African Crocodiles. *PLoS ONE* **9**, e94626.

Villiers A (1953) Contribution à l'étude du peuplement de la Mauritanie. Note sur la faune aquatique et ripicole de l'Adrar mauritanien. *Bulletin de l'Institut Fondamental d'Afrique Noire* **15**, 631-646.

Wagner P, Melville J, Wilms TM, Schmitz A (2011) Opening a box of cryptic taxa – the first review of the North African desert lizards in the Trapelus mutabilis Merrem, 1820 complex (Squamata: Agamidae) with descriptions of new taxa. *Zoological Journal of the Linnean Society* **163**, 884-912.

Wang C, Schroeder KB, Rosenberg NA (2012) A maximum-likelihood method to correct for allelic dropout in microsatellite data with no replicate genotypes. *Genetics* **192**, 651-669.

Ward D (2009) *The biology of deserts* Oxford University Press, New York.

Willing E-M, Dreyer C, van Oosterhout C (2012) Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE* **7**, e42649.

Appendix

Table S1. Information of the individuals used for the phylogenetic analyses. Each sub-basin encompasses the number of locations; belonging to basin and country, the number of individuals collected from each sub-basin and the lineage they belong to *Coptodon* A (CA), *Coptodon* B (CB), *Sarotherodon* A (SA) and *Sarotherdon* B (SB), with number of individuals (in brackets).

Sub-basin	Location (n)	Basin	Country	Indiv. (n)	Lineage
Tissint	3	Draa	Morocco	29	CA (7), SB (22)
El Mellah	1	Draa	Morocco	2	SB (2)
Zouwa	1	Draa	Morocco	7	CA (7)
Draa	2	Draa	Morocco	5	SB (5)
Coastal	1	Coastal	Southern Morocco	5	CB (5)
Coastal	2	Coastal	Mauritania	24	CB (13), SA (11)
Senegal	14	Senegal	Mauritania	30	CA (14), CB (2), SA (2), SB (12)
Rkiz	1	Senegal	Mauritania	1	SB (1)
Gorgol el Akhdar	9	Senegal	Mauritania	48	SB (48)
Garfa	3	Senegal	Mauritania	7	SB (7)
Niordé	2	Senegal	Mauritania	5	SB (5)
Touna	1	Senegal	Mauritania	1	CA (1)
Bounoukolé	1	Senegal	Mauritania	5	CA (5)
Karakoro	6	Senegal	Mauritania	28	CA (4), SB (24)
Niger	2	Korama	Niger	2	SB (2)
	49			**199**	

Table S2. List of the haplotypes based on mtDNA (16S+ND2) data. The number of haplotype (H), the number of samples (n), the lineage, sub-basin, country and clusters based on microsatellite data, the number of sample (between brackets).

H	n	Lineage	Sub-basin	Country	Cluster (n)
H1	7	*Sarotherodon* B	Gorgol el Akhdar, Grafa, Senegal	Mauritania	olive (6)
H2	1	*Sarotherodon* A	Senegal coastal	Mauritania	dark grey (1)
H3	28	*Sarotherodon* B	Senegal, Gorgol el Akhdar, Garfa, Niordé, Karakoro	Mauritania	olive (23), light yellow (3), orange (2)
H4	24	*Sarotherodon* B	Senegal, Gorgol el Akhdar, Garfa, Karakoro	Mauritania	olive (7), light yellow (15)
H5	5	*Sarotherodon* B	Senegal, Gorgol el Akhdar el Akhdar	Mauritania	olive (3)
H6	1	*Sarotherodon* B	Gorgol el Akhdar el Akhdar	Mauritania	olive (1)
H7	2	*Sarotherodon* B	Gorgol el Akhdar el Akhdar	Mauritania	olive (2)
H8	2	*Sarotherodon* B	Gorgol el Akhdar el Akhdar	Mauritania	olive (1)
H9	32	*Sarotherodon* B	Gorgol el Akhdar, Draa, Tissint, El Mellah	Mauritania, Morocco	olive (3), light grey (28)
H10	9	*Sarotherodon* A	Coastal, Senegal coastal	Mauritania	dark grey (8)
H11	2	*Sarotherodon* A	Coastal	Mauritania	dark grey (2)
H12	1	*Sarotherodon* A	Coastal	Mauritania	dark grey (1)
H13	1	*Sarotherodon* B	Niordé	Mauritania	olive (1)
H14	2	*Sarotherodon* B	Niordé	Mauritania	olive (2)
H15	4	*Sarotherodon* B	Gorgol el Akhdar	Mauritania	olive (4)
H16	1	*Sarotherodon* B	Senegal	Mauritania	olive (1)
H17	2	*Sarotherodon* B	Korama	Niger	olive (1), mix (1)
H18	1	*Sarotherodon* B	Senegal	Mauritania	orange (1)
H19	2	*Sarotherodon* B	Senegal, Gorgol el Akhdar	Mauritania	purple (2)
H20	4	*Sarotherodon* B	Senegal, Karakoro	Mauritania	purple (3)
H21	1	*Sarotherodon* B	Gorgol el Akhdar el Akhdar	Mauritania	olive (1)
H22	1	*Sarotherodon* B	Gorgol el Akhdar el Akhdar	Mauritania	olive (1)
H23	1	*Coptodon* A	Senegal coastal	Mauritania	-
H24	16	*Coptodon* A	Coastal, Senegal coastal	Mauritania, Southern Morocco	-
H25	3	Coptodon A	Coastal	Mauritania	-
H26	12	*Coptodon* B	Senegal, Bounoukolé	Mauritania	-
H27	9	*Coptodon* B	Senegal, Karakoro, Touna	Mauritania	-
H28	13	*Coptodon* B	Tissint, Zouwa	Morocco	-
H29	1	*Coptodon* B	Zouwa	Morocco	-
H30	1	*Coptodon* B	Bounoukolé	Mauritania	-

Table S3. List of taxa used for the phylogenetic analyses. Names of the species with GenBank accession number for each gene included in databases Figure 7 and 8.

Species	16SrRNA	ND2	1st Intron of S7
Etia nguti	GQ167966	GQ167777	GQ168091
Gobiocichla wonderi	GQ167967	GQ167778	GQ168092
Heterochromis multidens	GQ167968	GQ167779	GQ168093
Oreochromis niloticus	GQ167969	GQ167780	GQ168094
Alcolapia alcalica	GQ167970	GQ167781	GQ168095
Oreochromis tanganicae	GQ167971	GQ167782	GQ168096
Pelmatochromis buettikoferi	GQ167972	GQ167783	GQ168097
Pelmatochromis nigrofasciatus	GQ167973	GQ167784	GQ168098
Pterochromis congicus	GQ167974	GQ167785	GQ168099
Sarotherodon nigripinnis	GQ167976	GQ167787	GQ168101
Sarotherodon sp. aff. galilaeus "mudfeeder"	GQ167977	GQ167788	GQ168102
Stomatepia mariae	GQ167985	GQ167796	GQ168110
Tilapia discolor	GQ167990	GQ167801	GQ168115
Tilapia tholloni	GQ167993	GQ167804	GQ168118
Oreochromis andersoni	GQ167994	GQ167805	GQ168119
Tylochromis sp.	GQ167998	GQ167809	GQ168123
Sarotherodon mvogoi	GQ168000	GQ167811	GQ168125
Tristramella simonis	GQ168002	GQ167813	GQ168127
Steatocranus tinanti	GQ168006	GQ167817	GQ168131
Sarotherodon caudomarginatus	GQ168008	GQ167819	GQ168133
Tilapia dageti	GQ168010	GQ167821	GQ168135
Tilapia louka	GQ168011	GQ167822	GQ168136
Tilapia mariae	GQ168012	GQ167823	GQ168137
Tilapia ap. aff. rheophila "Samou"	GQ168014	GQ167825	GQ168139
Tilapia cf. nyongana "Dja"	GQ168016	GQ167827	GQ168141
Tilapia sp. aff. zillii "Kisangani"	GQ168018	GQ167829	GQ168143
Iranocichla hormuzensis	GQ168019	GQ167830	GQ168144
Tilapia zilli	GQ168025	GQ167836	GQ168150
Tilapia guineensis	GQ168026	GQ167837	GQ168151
Tilapia ruweti	JX910608	JX910825	JX910788
Coptodon zillii	JX910610	JX910881	JX910790
Sarotherodon galilaeus	JX910613	JX910869	JX910793
Orthochromis kalungwishiensis	JX910617	JX910861	JX910797
Coptodon cameruensis	JX910618	JX910877	JX910798
Pelmatolapia mariae	JX910620	JX910876	JX910800
Coptodon rendalli	JX910623	JX910889	JX910803
Coptodon zillii	JX910624	JX910878	JX910804
Sarotherodon sanagaensis	JX910625	JX910868	JX910805
Coptodon bakossiorum	JX910627	JX910885	JX910807
Heterotilapia buttikoferi	JX910628	JX910892	JX910808
Coelotilapia joka	JX910629	JX910895	JX910809
Coptodon zillii	JX910630	JX910879	JX910810
Coptodon bemini	JX910631	JX910888	JX910811
Coptodon bythobathes	JX910632	JX910886	JX910812
Coptodon zillii	JX910633	JX910882	JX910813

Coptodon rendalli	JX910634	JX910890	JX910814
Coptodon deckerti	JX910635	JX910883	JX910815
Coptodon snyderae	JX910636	JX910887	JX910816
Coptodon sp. "Cross"	JX910637	JX910884	JX910817
Sarotherodon lohbergeri	JX910638	JX910870	JX910818
Konia eisentrauti	JX910639	JX910871	JX910819
Coptodon zillii	JX910640	JX910880	JX910820

Table S4. List of the haplotypes based on nDNA sequencing data. The number of haplotype (H) and haplotype with detected recombination (HR*), the number of phased sequences (n), the lineage, sub-basin, country and clusters based on microsatellite, the number of sample is between brackets.

H	n	Lineage	Sub-basin	Country	Cluster (n)
H1	2	*Coptodon* B	Zouwa	Morocco	-
H2	1	*Coptodon* A	Coastal	Southern Morocco	-
H3	10	*Coptodon* A	Coastal, Senegal	Southern Morocco, Mauritania	-
H4	3	*Coptodon* A	Coastal	Mauritania, Southern Morocco	-
H5	12	*Sarotherodon* B	Draa, Tissint, El Mellah, Senegal, Gorgol	Morocco, Mauritania	orange (3) light grey (9)
H6	8	*Sarotherodon* B	Tissint, Senegal, Karakoro	Morocco, Mauritania	light yellow (1), orange (2), light grey (5)
H7	22	*Sarotherodon* B	Senegal coastal, Senegal, Gorgol el Akhdar, Garfa, Niordé, Karakoro, Korama	Mauritania, Niger	olive (20) , light yellow(1)
H8	13	*Sarotherodon* B	Senegal, Gorgol el Akhdar, Karakoro, Korama	Mauritania, Niger	olive (11), light yellow (2)
H9	8	*Sarotherodon* A	Coastal, Senegal	Mauritania	dark grey (9)
H10	3	*Sarotherodon* B	Senegal, Gorgol el Akhdar	Mauritania	purple (3)
H11	1	*Sarotherodon* A	Senegal coastal	Senegal	dark grey (1)
H12	1	*Sarotherodon* A	Senegal coastal	Senegal	dark grey (1)
H13	1	*Sarotherodon* B	Gorgol el Akhdar	Mauritania	orange (1)
H14	1	*Sarotherodon* B	Gorgol el Akhdar	Mauritania	purple (1)
HR1*	3	Coptodon B	Senegal, Karakoro	Mauritania	HR1*
HR2*	1	Coptodon B	Senegal	Mauritania	HR2*
HR3*	1	Sarotherodon B	Gorgol el Akhdar	Mauritania	HR3*
HR4*	1	Sarotherodon B	Gorgol el Akhdar	Mauritania	HR4*
HR5*	1	Sarotherodon B	Rkiz	Mauritania	HR5*
HR6*	1	Sarotherodon B	Rkiz	Mauritania	HR6*

* The haplotypes have been excluded from further analyses.

Table S5. **Partitions used for phylogenetic analyses.** The partitions and the best evolutionary models identified using Partition Finder.

Subset	Best Model	Subset partitions	Subset sites (bp)
1	GTR+I+G	16S	1-504
2	GTR+I+G	ND2_1	505-1239/3
3	HKY+I	ND2_2	506-1239/3
4	GTR+G	ND2_3	507-1239/3
5	HKY+G	S7	1-492

Table S6. Distribution of two lineages, *Sarotherodon* A and B, used in microsatellite analyses. Each sub-basin encompasses the number of locations; belonging to basin and country, and the number of individuals collected from each sub-basin.

Sub-basin	Location (n)	Basin	Country	Individuals (n)
Sarotherodon A				
Coastal	2	Coastal	Mauritania	10
Senegal	2	Senegal	Mauritania	2
Sarotherodon B				
Tissint	3	Draa	Morocco	21
El Mellah	1	Draa	Morocco	2
Draa	2	Draa	Morocco	5
Senegal	5	Senegal	Mauritania	12
Gorgol el Akhdar	9	Senegal	Mauritania	40
Garfa	3	Senegal	Mauritania	7
Niordé	2	Senegal	Mauritania	5
Karakoro	5	Senegal	Mauritania	22
Korama	2	Niger	Niger	2
	36			**128**

Table S7. Characteristics and genetic diversity of all microsatellite loci for 6 clusters (K=6) of both lineages *Sarotherodon* A and B. The allele size range (bp), number of alleles (N_a), observed (H_O) and expected (H_E) heterozygous.

	Sarotherodon B				*Sarotherodon* A				*Sarotherodon* B											
	Cluster 1 (n=28) Light grey				Cluster 2 (n=12) Dark grey				Cluster 3 (n=5) Purple				Cluster 4 (n=3) Orange				Cluster 5 (n=59) Olive				Cluster 6 (n=19) Light yellow			
Locus	Size range	N_a	H_O	H_E	Size range	N_a	H_O	H_E	Size range	N_a	H_O	H_E	Size range	N_a	H_O	H_E	Size range	N_a	H_O	H_E	Size range	N_a	H_O	H_E
Til1	227-226	4	0.536	0.662	230-250	4	0.750	0.659	230-250	5	0.800	0.844	222-246	3	0.333	0.600	234-266	6	0.695	0.671	242-258	4	0.421	0.368
Til2	192	1	-	-	180-182	2	0.091	0.091	188-190	2	0.200	0.200	208-220	5	0.667	0.933	178-212	13	0.847	0.856	190-212	5	0.722	0.684
Til13	148-176	4	0.571	0.627	178-180	2	0.083	0.083	140-156	2	1.000	1.000	140-148	3	0.000	0.800	120-188	17	0.780	0.897	120-172	6	0.737	0.735
Til8	128	1	-	-	116	1	-	-	110-158	7	1.000	0.964	116-140	4	1.000	0.867	125-155	11	0.860	0.847	140-146	3	0.316	0.317
Til19	196-216	6	0.786	0.729	208-268	11	0.909	0.922	192-216	6	1.000	0.889	188-236	5	0.667	0.933	164-224	12	0.881	0.853	184-212	5	0.684	0.674
Til20	138-140	2	0.357	0.468	-	-	-	-	130	1	-	-	138-140	2	0.333	0.333	142-148	3	0.368	0.508	142-148	3	0.526	0.649
Til6	131-151	5	0.571	0.643	105-135	7	0.917	0.830	105-111	4	0.600	0.711	125-139	4	0.667	0.867	115-149	14	0.825	0.815	133-139	4	0.684	0.539
Til21	286	1	-	-	268-304	5	0.750	0.703	260-270	3	0.600	0.644	274-298	4	0.667	0.867	268-294	9	0.576	0.746	272-294	4	0.737	0.738
Til22	160	1	-	-	150-156	3	0.333	0.540	164-168	2	0.200	0.200	154-182	4	0.667	0.867	1601-64	3	0.390	0.357	1601-64	2	0.263	0.235
Til23	104	1	-	-	104	1	-	-	104	1	-	-	104	1	-	-	110-114	3	0.068	0.067	112	1	-	-
Til24	230	1	-	-	230-232	2	0.167	0.290	232-234	2	0.200	0.200	230	1	-	-	230-234	3	0.254	0.312	232	1	-	-
Til25	206-216	4	0.321	0.366	166	1	-	-	164-166	2	0.600	0.467	194-216	5	1.000	0.933	174-200	2	0.017	0.017	174	1	-	-
Til30	111	1	-	-	111	1	-	-	111-113	2	0.200	0.200	111	1	-	-	111-121	4	0.559	0.509	117-119	2	0.211	0.193

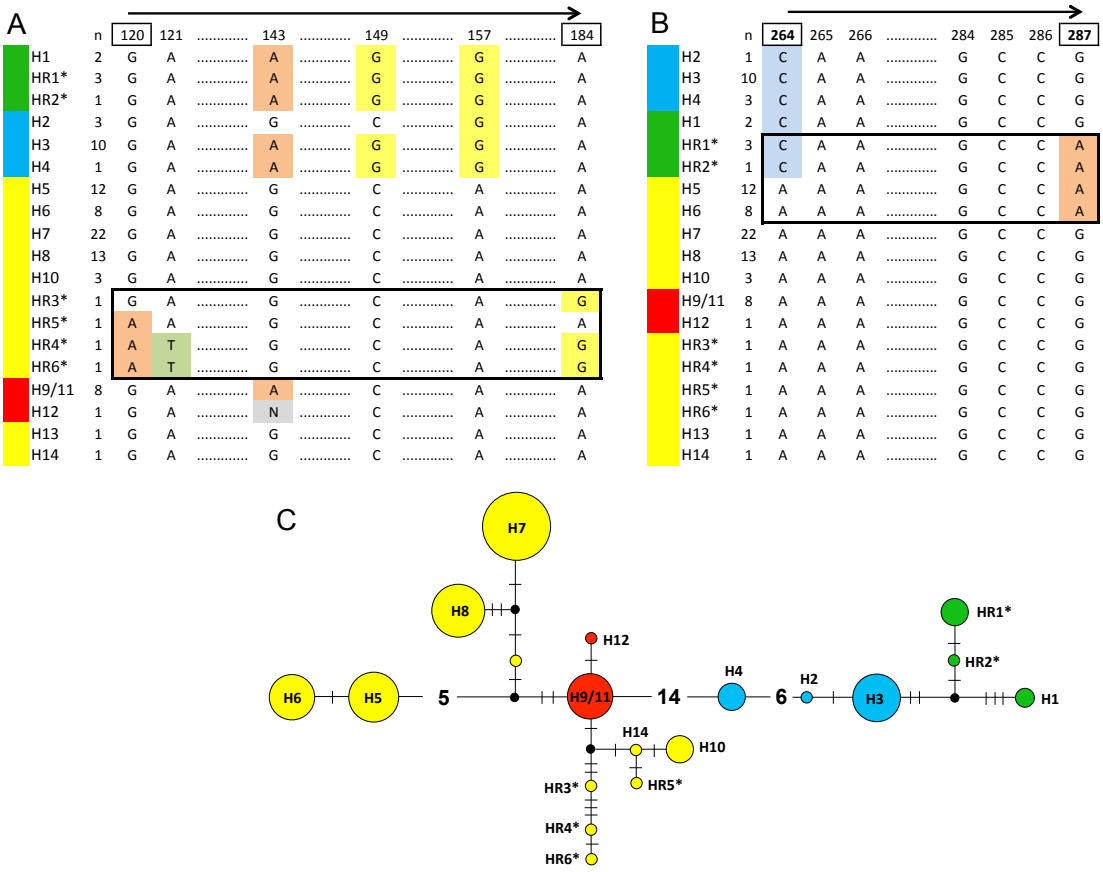


Figure S1. Recombination events and haplotype network. Recombination detected within *Sarotherodon B* (in yellow) in site (120, 184) and within (B) *Coptodon* B (in green) in site (264, 287). The colours near haplotypes represent the lineage: blue – *Coptodon* A, green – *Coptodon* B, red – *Sarotherodon* A and yellow – *Sarotherodon* B. The recombinant sequences are highlighted by black boxes and haplotype marked by *. (C) Haplotype network representing with all haplotypes. Circle size is proportional to sample frequency, black dots represent inferred missing haplotypes and bars/ numbers correspond to the observed mutations.

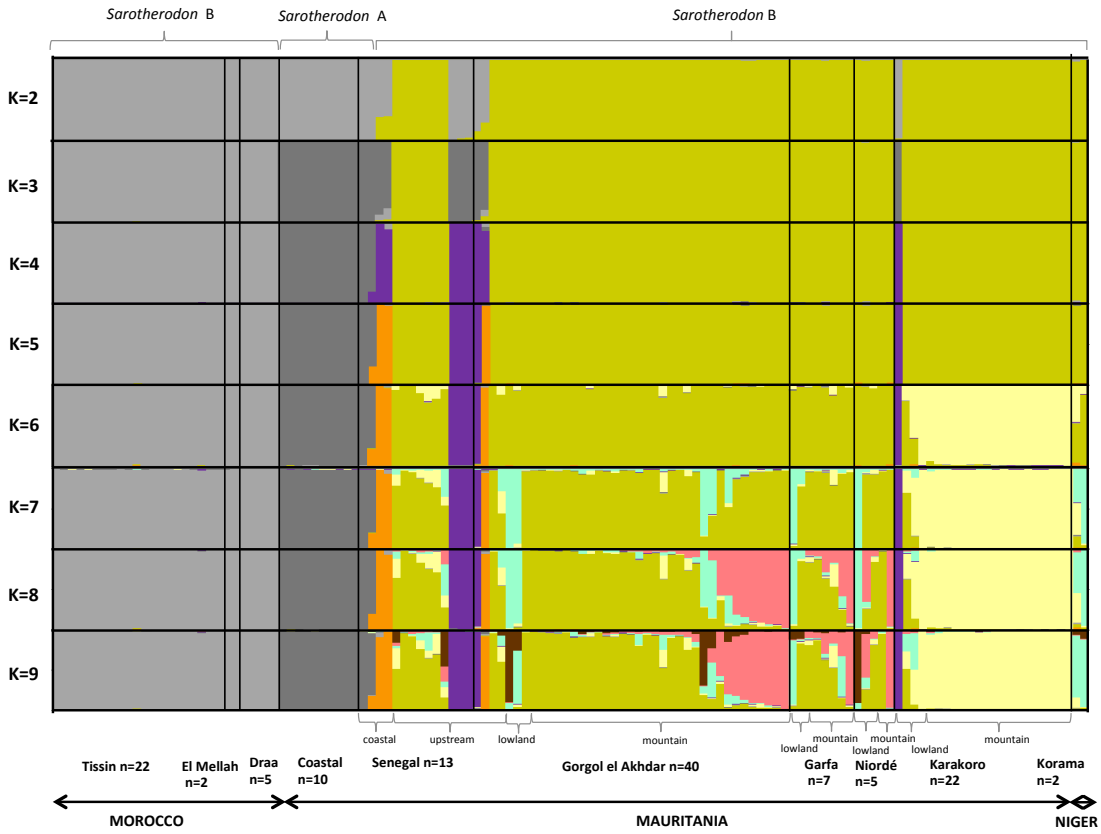


Figure S2. Population structure analyses of *Sarotherodon* A and B lineages across Morocco, Mauritania and Niger. The bar plots are presented since K=2, which is the best K according to Evanno's method, until K=8, which is the best according ln Pr(X|K) method. With one more bar plot with K=9, to visualize the further clustering. Each vertical line represents an individual and each colour represents a cluster. Individuals are grouped by basins or sub-basins, which are divided by black vertical lines, and with their names in the bottom of plot bar, with the number of samples shown into brackets. Each Mauritanian sub-basin is divided by topology shape using braces.

Primer Note

Development and characterization of 13 microsatellite loci of *Sarotherodon* genera from Mauritanian inland

Abstract

Thirteen polymorphic microsatellite loci were developed for Tilapias from Mauritania inland, using 454 GS-FLX Titanium pyrosequencing platform. These markers were characterized in 38 individuals from Gorgol el Akhdar sub-basin in Mauritania, ranging from 2 to 17 alleles per loci, with an average of 7. The expected heterozygosity ranged from 2.6% to 84.8% with an average of 56.5%. The newly developed microsatellite loci can be used to describe the population structure in *Sarotherodon* genera in Mauritania.

Sarotherodon species belong to African Tilapias, which are also distributed in Levant (Beveridge & McAndrew 2000) . Tilapia name hold many species, but is more commonly used for three genera *Sarotherodon*, *Tilapia* (*Coptodon)* and *Oreochromis* (Trewavas 1983). They belong to the most reach vertebrate family, including until 3, 000 species, Cichlidea (cichlids) family (Kocher 2004). Tilapias together with other cichlids are important as food for human, aquaculture and have big interest of scientist of their complex in systematics, taxonomy, ecology, morphological variation and behavour. The East African cichlids are known as impressive example of adaptive radiation (Beveridge & McAndrew 2000). While the expeditions to Sahara since the beginning of the 20th century increase the knowledge about relict fish species in isolated water patches in those areas, including cichlid species (Lévêque 1990), but still leaving a gap of knowledge about not sampled areas and not described species. These relict taxa are interesting to study to understand better the evolutionary processes and mechanism, which prevent them to exhibit small habitat spots and survive per harsh climatic conditions. Here, our interest was to develop and characterize a set of microsatellite markers useful for studying *Sarotherodon* genera fishes at a population structure level, mainly across Mauritanian fresh water bodies.

Thirteen individuals, from the *Sarotherodon* genera (Lineage *Sarotherodn* A, see for details in Results), were collected at different locations in Mauritania, including water bodies from rocky and lowland areas, and Senegal River. The genomic DNA (gDNA) was extracted from fins or muscle tissue using EasySpin® Genomic DNA Tissue Kit (Citomed), and was quantified in a fluorometer (Perkin-Elmer) using Quant-iT™ PicoGreen®dsDNA Assay Kit (Invitrogen). The microsatellite loci were isolated for the *Sarotherodon* lineage at the 454 GS-FLX Titanium pyrosequencing platform (GS FLX®, Roche Diagnostics) of Genoscreen, France (www.genoscreen.fr). Libraries were built from pooled samples of 13 individuals, totaling 3 µg of genomic DNA, and enriched for 32 motifs and compliments of motifs. Individuals were collected microsatellite library contained 261 markers with 32 different motifs, recommended primer sets and multiple FASTA file with sequences. Thirty primer sets were selected and distributed in three multiplexes, based on the following criteria: nucleotide motif up to tetra, number repeats higher than 10, higher primer quality (given by Genoscreen), lower PCR product size and lower difference in primer melting temperature. Each group of primers was checked in AUTODIMER software to avoid primer-dimer and hairpin interaction in short DNA oligomers (Vallone & Butler 2004). Forward primers were tagged with tails to attach the fluorescence dyes (FAM, VIC, PET and NED; Applied Biosystems), whereas reverse primers were tagged with "pig-tails" (GTTT-sequence tag), to avoid problems caused by plus-A artefacts (Brownstein *et al.* 1996). Thirty-eight samples were collected at the Gorgol el Akhdar sub-basin, Mauritania. First, PCR reactions were performed for each primer pair set in order to confirm the expected PCR product size and polymorphism. PCR products were visualized in a 2% agarose gel and were screened on a 3130*xl* Genetic Analyzer (Applied Biosystems). Amplifications were performed on a final reaction volume of 10 µl, containing 5 µl of Qiagen© Multiplex PCR Kit Master Mix (Qiagen), 0.04 µM of forward primer, 0.4 µM of reverser primer, 0.4 µM indicated florescence label (Table 1) 3.16 µl ddH_2O and approximately 10 ng of gDNA. PCR cycles consisted of an initial denaturation at 95 °C for 15 min, following by 9 cycles of denaturation at 95 °C (30 s), annealing at 56 °C (1 min 30 s) with decrees in each following cycle by 0.5 °C, extension at 72 °C (45 s); 31 cycles at 95 °C (30 s), annealing at 52 °C (1 min), extension at 72 °C (45 s) and final extension step at 60 °C (10 min). Markers showing stutter bands and no polymorphism were discarded. Thirteen, of the 30 microsatellite loci tested, were grouped into two multiplexes and one singleplex (*Til6*). The marker *Til6* was amplified using previously described PCR reaction conditions. The two multiplexes were amplified in a final reaction volume of 10 µl, containing 5 µl of Qiagen© Multiplex PCR Kit Master Mix, 1

Table 1. Characteristics and genetic variation of 13 microsatellites in *Sarotherodon B*: locus name, singleplex/ multiplex, repeat motif, primer sequences, fluorescent label M13-tail, multiplex, primer concentration (μM), number of individuals analysed (*n*), allele size range (bp), number of alleles (N_a), observed (H_O) and expected (H_E) heterozygosity, Hardy-Weinberg *P*-value and GenBank accession number.

Locus	Repeat motif	Fluor. label	Primer sequences (5´-3´)	Primer (μM)	*n*	Size range (bp)	N_a	H_O	H_E	H-W (*P*-value)	GenBank accession no.
Til6	$(AC)_{13}$	VIC	F: CCCTCCTGCTTAAGTCAACAC R: TAAGGGTTCCCGACTCTCCT	0.4	37	115 - 149	13	0.838	0.817	0.392	
Mix1											
Tlil1	$(ATAG)_{12}$	6-FAM	F: TGCTAAGAACAAAGTCTTGGGA R: TCATCATGCTGCGGTAACAC	0.4	38	242 - 266	5	0.763	0.691	0.851	
Til2	$(CA)_{14}$	6-FAM	F: TGGAACACTTTGGAGGATCG R: CGTTCCTGGATCAAACCG	0.4	38	178 - 212	11	0.816	0.833	0.235	
Til13	$(CCAT)_{10}$	6-FAM	F: GAGCTTGGCCATTTAGGAGA R: TGTTGGAACATAAGGATAAGTGGA	0.8	38	120 - 188	17	0.816	0.888	0.014	
Til8	$(CAA)_{15}$	NED	F: CCCACAAAAGAATTAAAGCTAAAGA R: TGAGTCTAACATTTGGTCTTTGAA	1.7	36	125 - 152	9	0.833	0.802	0.952	
Til19	$(TCTA)_{15}$	PET	F: ATCAGGGTCGTGACTTCTGC R: CCAAATGAGCAGGATGATTG	1	38	176 - 216	9	0.921	0.838	0.183	
Til20	$(CA)_{13}$	PET	F: CAGCTTCCCGTGCTCATTAT R: GCAGCACCTGGAAGATAAAA	1.8	36	142 - 148	3	0.306	0.406	0.039	
Mix2											
Til21	$(TGGA)_9$	6-FAM	F: GGTCCAAAGACATAATTAGTGGG R: TGTGTTGGGTAAGTATCTTCTAGTTCA	0.4	38	268 - 294	9	0.553	0.713	0.076	
Til22	$(GT)_{12}$	6-FAM	F: AAACCAATAAATACAACCCACTG R: TTAACACAGTTTCTTATACGACAGAGA	0.4	38	160 - 164	3	0.447	0.400	0.404	
Til23	$(CA)_{10}$	6-FAM	F: TTAGTCAGCAGCCAACACCC R: GACCTCCTCGTGAAGTGCAT	0.35	38	110 - 114	3	0.053	0.052	1	
Til24	$(AC)_8$	VIC	F: TGAACGTGCAGCAGAGTCAC R: GGAGCTCTTTAGGTGGTTGG	0.4	38	230 - 234	3	0.316	0.326	0.372	
Til25	$(CA)_{14}$	VIC	F: GGGGCCTCAGTCTAGGAAAG R: GGAGTGGGGACGTGCATA	0.5	38	174, 200	2	0.026	0.026	-	
Til30	$(TG)_{11}$	PET	F: CTGACAGCAAGAGCCTCAAA R: GAAATAGGTTCTGGCTCAAAGTT	0.6	38	111 - 121	4	0.658	0.545	0.587	

μM of primer Mix (see Table 1 for details), 3 μl ddH_2O and approximately 1 μl of genomic DNA. Similar thermocycling conditions were used with the final extinction increased to 30 min. PCRs were performed on a BioRad C1000 Thermocycler. PCR products were visualized in a 2% of agarose gel and fragment sizes were determined on a 3130*xl* Genetic Analyzer (AB), using using GeneScan™ 500 LIZ® Size Standard (Applied Biosystems). The allele calling was performed in GeneMapper v4.1 (Applied Biosystems). Microsatellite toolkit was used for error and duplication detection in data (Park 2001). The estimates of allelic diversity, observed and expected heterozygosities, were calculated using GenALEx (Peakall & Smouse 2006). Tests for deviation from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium were performed in GENEPOP (Raymond & Rousset 1995) and the statistically significance for multiple test was established using the Bonferroni correction.

The observed number of alleles ranged from two (*Til25*) to 17 (*Til13*), with an average of seven alleles per locus (Table 1). The expected heterozygosity ranged from 2.6% (*Til25*) to 88.8% (*Til13*), and the observed heterozygosity ranged from 2.6% (*Til25*) to 92,1% (*Til19*), with the average of 56.5% and 56.4%, respectively. All loci showed no deviation from Hardy-Weinberg equilibrium and any evidence of linkage disequilibrium were found, after Bonferroni correction. Although only six of the 13 loci (*Til2*, *Til3*, *Til8*, *Til19*, *Til6*, *Til21*) were found to be highly polymorphic, however all maker set was found to have enough power to distinguish more genetically distant populations (see the results of population genetics). These markers will be used to study the genetic structure and gene flow among Mauritanian populations of the *Sarotherodon* lineage.

References

Beveridge MC, McAndrew BJ (2000) *Tilapias: biology and exploitation* Springer.

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* **20**, 1004-1006, 1008-1010.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics* **5**, 288-298.

Lévêque C (1990) Relict tropical fish fauna in Central Sahara. *Ichthyol. Explor. Freshwaters* **1**, 39-48.

Park S (2001) The Excel Microsatellite Toolkit (version 3.1). Animal Genomics Laboratory, UCD, Ireland.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of heredity* **86**, 248-249.

Trewavas E (1983) Tilapiine fishes of the genera Sarotherodon, Oreochromis and Danakilia.

Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* **37**, 226-231.