University of Natural Resources and Life Sciences, Vienna
Department of Integrative Biology and Biodiversity Research

MASTERARBEIT/MASTER THESIS

Titel der Masterarbeit/Title of the Master Thesis

"Molecular diversity of Burkina Faso's freshwater fish in the Mouhoun and Comoé catchment assessed with DNA barcoding."

verfasst von/submitted by

Nikolaus Schobesberger, BSc

Angestrebter akademischer Grad / for the degree of

Master of Science (MSc)

Studienkennzahl / degree programme code:	H 066 448
Studienrichtung / degree programme:	Applied Limnology
Betreut von / Supervisor:	**Univ.Prof. Dipl.-Biol. Dr.rer.nat Harald Meimberg**
Mitbetreut von / Co-Supervisor	**Priv. -Doz. Dipl.-Ing. Dr. Andreas Melcher,** **Paul Meulenbroek, MSc. Dr.**

Eidesstattliche Erklärung

Ich erkläre eidesstattlich, dass ich die Arbeit selbständig angefertigt, keine anderen als die angegebenen Hilfsmittel benutzt und alle aus ungedruckten Quellen, gedruckter Literatur oder aus dem Internet im Wortlaut oder im wesentlichen Inhalt übernommenen Formulierungen und Konzepte gemäß den Richtlinien wissenschaftlicher Arbeiten zitiert, durch Fußnoten gekennzeichnet bzw. mit genauer Quellenangabe kenntlich gemacht habe.

_____ _____

Ort, Datum Unterschrift

Abstract

Burkina Faso is a landlocked country in the soudano-sahelian zone, where fish are an important part of the daily protein intake of its population. Fish hauls however have plateaued in the last decades and fish sizes have dramatically decreased as overfishing, loss of habitat and fragmentation of water bodies have become prevalent.

A detailed knowledge of fish species is indispensable for a responsible and successful monitoring of water bodies and its fish populations. Morphological identification has been used in the past to describe the ichthyofauna of this country, but it does not recognise cryptic species that some genera may hold. Here, DNA barcoding is considered to be an efficient tool to reveal divergences in species that are difficult to distinguish morphologically. The SUSFISH project aims to build capacities to study, monitor and manage sustainable fisheries in Burkina Faso. In the line of this project, this study will employ molecular tools to shed light on taxonomic inconsistencies in Burkina Faso's freshwater fish which are prevalent in some genera like *Synodontis*. The objective of this thesis is to examine the barcoding region with Illumina Sequencing and Sanger Sequencing in order to provide the basis for building a DNA barcode reference library for Burkina Faso's freshwater fish. 411 samples were obtained and sequenced. The sequences were used to construct haplotypes. These were then used to produce a neighbour joining phylogenetic tree. The results suggest 4 species that harbour cryptic species within their lineages.

Acknowledgements

I want to thank the OeAD for funding the APPEAR program with its SUSFISH project in which this study took place, the Centre for International Relations (ZIB) from BOKU University for its scholarship for short-term scientific research work abroad (KUWI) and the Burkinabe SUSFISH partners who organized the trip.

I would like to thank my thesis' supervisor Dr. Harald Meimberg for his support during the process of writing this thesis and his patience. I could always count on his advice when I needed it. I would also like to thank Dr. Andreas Melcher for the opportunity to participate in the SUSFISH project and Dr. Paul Meulenbroek for his support during fieldwork and afterwards.

I want to especially thank Dr. Manuel Curto and Eva Dornstauder-Schrammel, DI for their support during laboratory work and guidance. Without their patience I would not have been able to conduct my work properly and their willingness to make time for me even in stressful circumstances is greatly appreciated.

Furthermore, I would like to thank all my colleagues from Ouagadougou University and BOKU University who went into the field with me. Without them none of this would have been possible.

Last but not least I want to thank my family, my friends and Susi, all of whom supported me throughout my studies. Thank you!

Index

1. Introduction

1.1. Burkina Faso and its fish fauna

Ecosystems are under immense pressure worldwide. The decline in biodiversity is especially prevalent in aquatic ecosystems and even more so in freshwater systems with a decline of 50% in biodiversity between 1970 and 2000 as documented by the Living Planet Index (WWF Living Planet Report 2014). Their greater vulnerability can be attributed to them being closer related and connected to human undertakings. Man made changes to habitat, like the damming of rivers, the introduction of pollutants and abstraction of resources are some of the factors contributing to the issues. 48% of rivers worldwide are moderately to severely impacted by flow regulation or fragmentation (Grill et al. 2015).

One of the countries severely affected by these man-made changes is the western African country Burkina Faso. The topography of the land locked country is flat with the average altitude ranging from 250 to 350 m. Burkina Faso has a tropical climate with three main climatic zones, one of which is located in the North with temperatures up to 45°C, a Soudano-sahelian zone in the centre with intermediary aridity and a Soudano-guinean zone in the South which has precipitation of up to 1000 mm. Its geography can be divided into three basins, the most prominent of which is the Volta-basin. The Mouhoun, the Nakambé as well as the Pendjari are all draining into this basin (Mason & Knight, 2011).

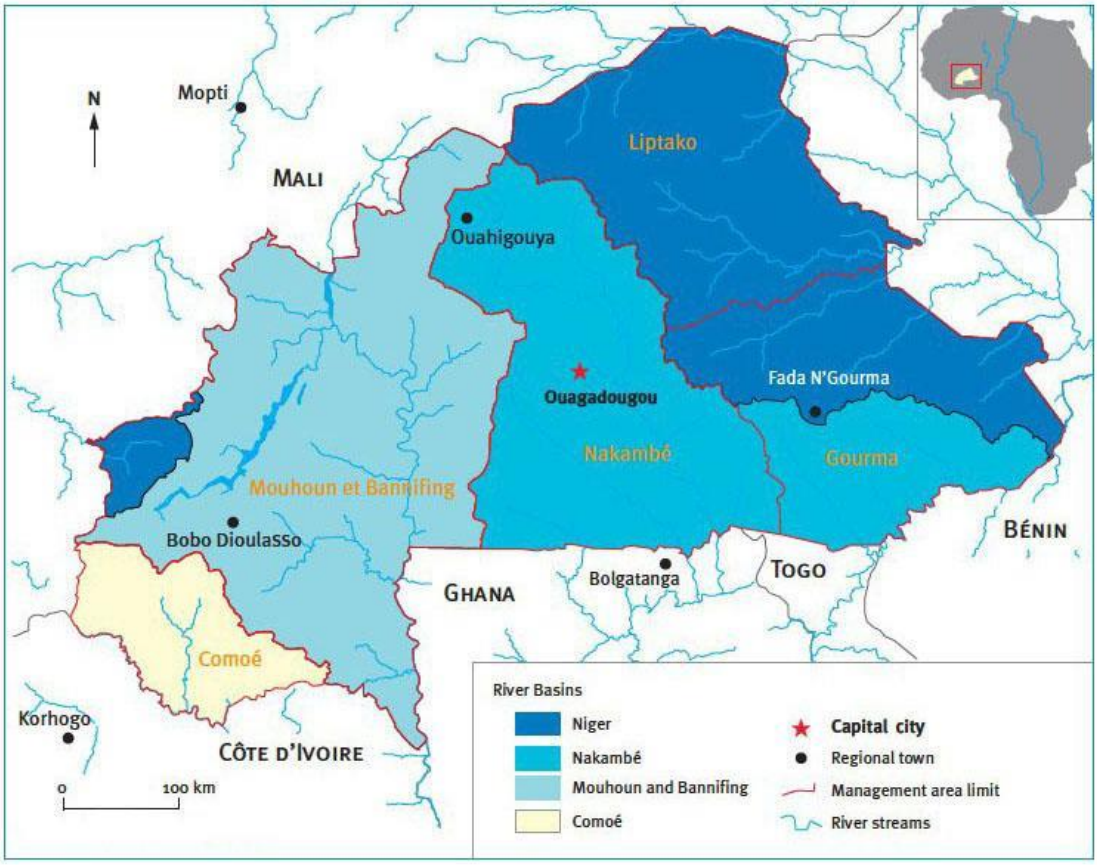

Figure 1: Map of Burkina Faso's rivers and river basins. The red lines are Management area limits and are congruent with the river basin divides, indicating the importance of water resource management. Source: Venot et al. 2014.

It is one of the world's poorest countries and is listed as the eight least developed by the human development index (UNDP 2019). The country's economy struggles on the one hand with natural challenges like severe droughts as well as floods and on the other hand with economic difficulties. 43,7 % of the 20,1 million people were below poverty level in 2014 (The World Bank Group 2019) with a literacy rate of 41 % among adults in 2018, which has doubled since 2003 (uis.unesco.org).

Since around 80% of the people of Burkina Faso are farmers, they depend on a constant supply of fresh water (Mason & Knight, 2011). So, to overcome the dry season and water scarcity, more than 1400 barrages were built from the 1950 onward, varying in size from 1 to 25 000 ha (Melcher et al. 2011, Ouedraogo, 2010). This, along with policies, which incentivised Burkinabe people to take up fishing as a profession (Bouda, 2002), led to a 15-fold increase of fisheries landings since 1950. In Burkina Faso, fishery is an important source

of protein as well as income. Fishing hauls however, have reached a plateau as the same time as fish size dramatically declined, severely cutting the productivity of fisheries (Melcher et al., 2014).

The land use along the reservoirs is often heavily affecting the water therein. Crop fields along the reservoirs are treated with chemicals and irrigated with water from the barrages leading to various problems downstream like unregulated water abstraction and pollution with agricultural substances (De Fraiture, 2014). While tolerated, this practice is unauthorized and just one of many human influences contributing to the degradation of Burkina Faso's freshwater resources. To manage freshwater resources and fisheries, some projects have been undertaken, one of which is the SUSFISH project.

The purpose of this study is to broaden the understanding of the river's ecosystem in line with the SUSFISH project. This project takes a multifaceted approach in river and fisheries management with several research projects in order to equip decision makers with competent tools to drive the socioeconomic process forward and in the right direction. Nowadays, a detailed knowledge of the DNA of fish is indispensable for a sophisticated understanding and handling of the country's resource that is fishery. In order to successfully operate a standardised monitoring and assessing of fish stocks, molecular determination is necessary (Nwani et al. 2011).

In the past, West African freshwater fish have been described using morphological features. While this method can be used reliably to some extent for distinguishing species, it does not take into account the cryptic diversity a genus or even a single species may show on a molecular level. DNA barcoding has shown to be an efficient tool for showing divergence among species that may look morphologically indistinguishable (Knebelsberger et al. 2014). It can be used to detect cryptic diversity in species or reveal taxonomic inconsistencies (Collins, Cruickshank, 2013), which are prevalent in Burkina Faso's freshwater fish (Meulenbroek, P. 2013). This may be especially true for the genus *Synodontis,* where interspecies differences are thought to be prevalent. Even experts in Burkina Faso are not sure in regard to species number and morphologically differentiating characteristics. Previous studies are split over the number of *Synodontis* species in Burkina Faso and neighbouring countries. While some

studies identified up to nine species morphologically (Mano, 2016), fishbase.org (version 12/2020) lists 13 species of *Synodontis* in the country. It is thought that hybridization between species further aggravates these difficulties. In the course of creating a DNA barcode reference library for the freshwater fish of Burkina Faso, this question may be addressed to get an overview of the genus *Synodontis* both morphologically and genetically.

These molecular assessment methods can be powerful to help the determination and description of species (Meier et al., 2006). The barcoding method has been proven to be highly successful in the identification of fish species, both marine and freshwater fish, with a success rate between 80% and even up to 100% (Pereira et al., 2013).

1.2. Molecular Methods

1.2.1. Introduction to DNA barcoding

The nucleotide sequence of any given organism contains information about its function and evolutionary history. That is why, since the 1970, genome sequencing has been a continuous field of research and new tools and technologies have been developed since then. Specifically, the COI region of the mitochondrial genome was proposed by Hebert et al. in 2003. Focusing on a single gene sequence to discriminate between the abundance of animal species, the 650 base pair fragments on the 5' end of the cytochrome c oxidase subunit I (COI) was chosen. This was done to introduce standardisation to previous different protein-based and DNA-based tools for molecular identification of, among others, fish species in Africa (Waters, Cambray 1997, Wishart et al., 2006, Swarz et al. 2008). The premise of DNA barcoding is the circumstance that the barcode sequence divergence between species is greater than the divergence within species (Hebert et al. 2003). This makes the COI region an effective marker for the identification of species and meaningful tool in their discovery. Since its conception, DNA barcoding has become a highly used, cost effective, fast and widely applicable method of species identification. The barcodes obtained by these studies are assigned to their respective species and specimens and uploaded to an online library, the barcode of life data system (BOLD; boldsystems.org; Ratnasingham, Hebert, 2007) where more than 1 million sequences are stored in combination with more than 94 000 species. This public online library supports the collection, storage, analysis and publication of DNA barcodes.

Since the distribution and diversity of freshwater fish in Burkina Faso is not known for sure to this day, one of the aims of this study is to determine if DNA barcoding can be used as an effective tool in aiding the species identification of freshwater fish in this country.

Two methods of sequencing were used in this study and are described in the following.

1.2.2. Sanger Sequencing

This method is based on the principle of discontinuation of a nucleotide sequence at a specific location. After a primer is attached to a single stranded DNA (ssDNA) fragment, it is complementarily extended until a certain deoxynucleotide (dADP, dTTP, dGTP, dCTP) base is reached. In its stead, a dideoxy nucleotide gets implemented. From this point on, the strand cannot be continued. On every location, where an Adenine should be, a copy of the strand ends. Likewise, this reaction takes place for the other three bases. In this way, a compilation of strands with certain length is made, which are visualized in conventional gel-electrophoresis. The sequence of bands in the gel corresponds to the sequence of nucleotides in the DNA fragment (Sanger et al 1977).

1.2.3. Next generation sequencing and Illumina Sequencing:

The next generation sequencing (NGS) Illumina is currently the dominant sequencing method on the market. Having evolved in the last years, this method allows for a significantly higher throughput than traditional sanger sequencing (Schirmer et al., 2016). Before Illumina sequencing can be done, a DNA library needs to be constructed in which DNA fragments that wish to be identified need to be equipped with adapters that contain index sequences. With an index PCR, these adapters are added and amplified. On a flow cell, these adapters bind to complementary oligonucleotides and are sequenced simultaneously in a row of sequencing reactions (Mardis, 2013).

To have a detectable signal for sequencing, another amplification is needed in a step called bridge amplification (Mardis, 2017). Here, the fragments are amplified *in situ* on the flow cell by building a bridge. Denaturation occurs with the original strand, where it splits from the adapter. After annealing, the steps of amplification and denaturation repeat until the flow cell is covered with DNA strands. Illumina sequencing employs a process called sequencing-by-synthesis (SBS) reaction. In this reaction, fluorescent nucleotides, which are modified in the

3’OH position, are added base-by-base. A laser illuminates the fluorescent nucleotides (A, T, G, C,) which give off a detectable signal, complimentary to the desired sequence (Bentley et al., 2008). Then, the 3’OH region of the nucleotides is chemically customized, and the process is repeated between 70 and 300 times to create reads of the same length. With the use of a

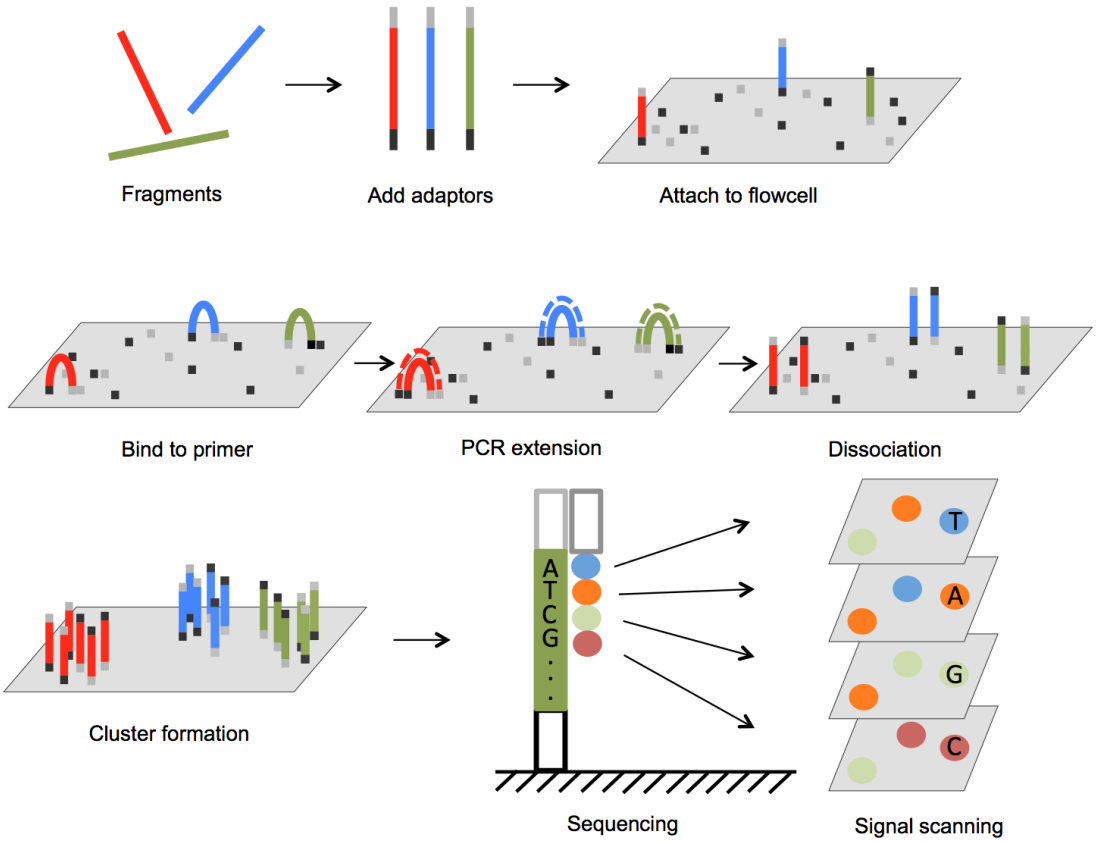

Figure 2: Simplified illustration of the Illumina sequencing steps. After adapters are attached with PCR, the fragments attach to the flow cell and bind to primers in the cell. Bridge amplification takes place. Cluster formation and excitation of modified nucleotides by laser. Optic scanning detects the signal and translates it into sequence. Source: Lu et al, 2015

reverse primer, a second read direction for each sequence is created (Mardis, 2013).

1.3. Aim of this study and overall objectives.

This study aims to get an overview of Burkina Faso’s freshwater fish diversity on a molecular level and provide the first steps in building a barcode reference library for utilization in the assessment of biodiversity and its conservation for the country. To achieve this, both Illumina sequencing and Sanger sequencing will be used, employing amongst others the genetic marker for the barcoding region. Special emphasis will be placed on the genus Synodontis with the intention to improve upon the existing uncertainties regarding the number species

this genus holds. An additional focus will be placed on the genetic makeup of species by comparing differences in the species' haplotypes between two river systems.

Overall objectives

- Assess Burkina Faso's molecular freshwater fish diversity by analysing DNA barcodes of morphologically identified material.
- Provide a first DNA Barcode Reference Library of Burkina Faso's Freshwater Fish

Specific Objectives

- Illustration of Burkina Faso's molecular freshwater fish diversity
- Analyse cryptic diversity within species, especially the genus *Synodontis*
- Comparison of selected species for different catchments/rivers

Research Questions

The fragmented nature of water bodies and their spatial and temporal discontinuity are causes for a restricted gene flow and can therefore lead to genetic divergence due to isolation (Decru et al. 2016).

(1) What is the actual biodiversity of fish species, especially the genus Synodontis, in Burkina Faso?

(2) What are the genetic distances of selected species between river catchments?

In order to answer these questions, the hypotheses are thus:

(i) There is a greater biodiversity of fish in Burkina Faso than previously thought, especially within the genus *Synodontis.*

(ii) Fish of the same species from varying catchments will have a different genetic makeup.

2. Material and Methods

2.1. Sampling

2.1.1. Sampling sites

Sampling took place in January of 2018, and in the first week of February 2018. Selection of sampling sites was done by local experts, supervisors, local fishermen and the University of Ouagadougou. Selection criteria were water availability, accessibility, security of the region, travelling costs as well as whether it was compatible with the research projects that were undertaken by accompanying colleagues. Last minute changes to the timetable and sampling sites had to be made due to various hindering factors like faulty equipment, sickness of the researcher and bureaucratic inconveniences. Habitats fished were small brooks, rivers, lakes and barrages. Sampling began in the Mouhoun catchment near Bobo Dioulasso, resumed in the Comé catchment around the small town of Banfora and was concluded around the village of Boura, where mainly the Mouhoun river was sampled. Sampling site coordinates are disclosed in the appendix 3 (table 9).

2.1.2 Sampling Methods

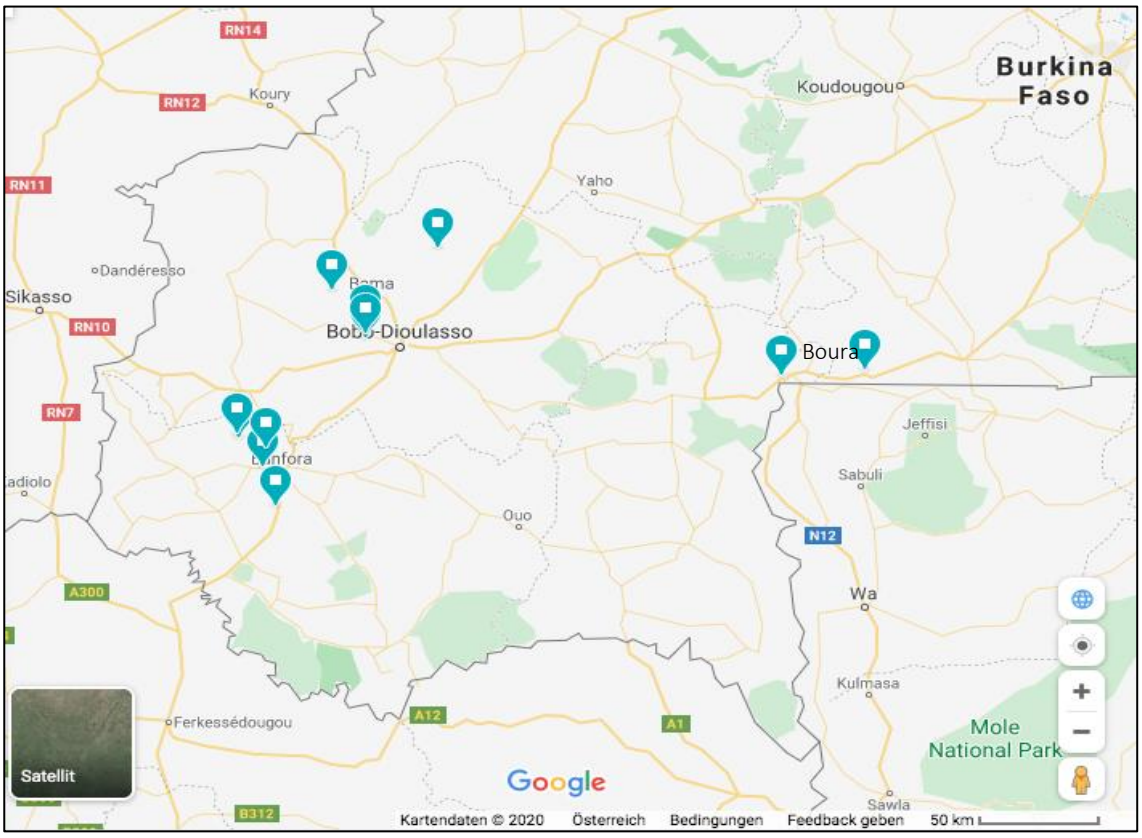

Figure 3.: Map of sampling sites. Source: Google Maps 2020

Figure 4.: Entering the muddy waters of a barrage with a throwing net on an overcast morning. Source: Hundscheid 2018

Figure 5.: After fishing, hunting for a photo opportunity of Hippos on Hippo Lake. Source: Hundscheid 2018

Figure 6.: Bargaining with local fishermen over selected specimens. Source: Hundscheid 2018

Figure 7.: Inspecting the catch of the day with local fishermen. Source: Hundscheid 2018

By the expertise of a local fisherman, cast nets, gill nets and dragging nets were used to catch fish. Mesh size of the traditional cast nets was 2.5 cm, for the GN it was either 50 mm or 10 mm. For dragging nets, a mosquito net was used. A few fish were also caught by the use of electric fishing with a backpack generator LTH60-IIH from Hans Grassl but it broke down after one use.

All samples were stored in ethanol, which was renewed after a few days to prevent degradation of genetic material. Specimens were identified by Dr. Mano Komandan and with the use of Paugy's 'The Fresh and Brackish Water Fish of West Africa', 2004.
From each sample, the left pectoral fin was used either partly or completely, depending on size. For smaller fish (<5cm) part of dorso-lateral muscle tissue was additionally used to compensate for small fins. Every piece of tissue was washed with 70 % ethanol after collection and dried before processing. DNA was then extracted using a standard DNA extraction protocol for metazoan cells/DNA.

2.2. DNA isolation and amplification

2.2.1. DNA isolation

The tissue was degraded using 300 µl of lysis buffer and 10 µl of Proteinase K (concentration 10 mg/ml) and a heating block, which was running for at least 3 hours at 56°C and 1000 rpm or until the tissue was dissolved. Afterwards 10 µl of RNase (10 mg/ml) was added and again ran for 20 minutes on the heating block at 37 °C and 1000 rpm. 75 µl of 3 M KOAc was added and the samples were put into -20°C for 20 minutes. Afterwards the samples were rotated in a centrifuge as follows: 1000 rpm for 1 minute, 2000 rpm for 1 minute, 4000 rpm for 1 minute, 8000 rpm for 1 minute, 11000 rpm for 7 minutes.
360 µl of supernatant was transferred to 2 ml deepwell plates and 540 µl of binding buffer was added and mixed 5 times with the pipette. 450 µl of the mixture was transferred to an EconoSpin™ Column plate (Epoch Life Science, USA) which was centrifuged with a swing out rotor as follows: 500 rpm for 2 minutes, 4000 rpm for 5 minutes and 6000 rpm for 5 minutes. The remaining 450 µl were transferred to the EconoSpin™ plates and centrifuged at 2000 rpm for 15 seconds and 4000 rpm for 1 minute.

Afterwards the EconoSpin™ plates were washed two times with 600 µl of 70 % Ethanol and each time centrifuged at 4000 rpm. After one final centrifugation of 6000 rpm for 1 minute, 4 elutions were made.

A 10 mM Tris Elution Buffer with a pH of 8 was added, left for incubation at room temperature for up to three minutes and each time centrifuged at 6000 rpm for 5 minutes. The volumes were 50 µl, 50 µl, 80 µl and 50 µl.

2.2.2. DNA and primer amplification

For each plate and elution, a gel was prepared with 0.8% agarose (Biozym Scientific GmbH, Germany) stained with Ethidium bromide and the samples that were selected for quality check and dyed with a ratio of 1:4 of DNA to blue dye and a Lambda DNA/HindIII ladder was added (Thermo Fisher Scientific, USA). These gels were run in a Gel electrophoresis bath for up to 20 minutes at 80 Volts and then checked with a Gel iX20 Imager.

Afterwards, Primers (VF2_t1, VR1_t1, Ward et al. 2005) for the Polymerase Chain Reaction amplification (PCR) (Meimberg et al., 2016) were determined using the Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1997). These Primers were attached and amplified using a PCR with the following PCR protocol: step 1: 15 minutes at 95°C denaturation, step 2: 95°C for 30 sec, step 3: 55°C annealing for 1 min, step 4: 72°C elongation for 1 min. 30 repeats for the steps 2 – 4, step 5: 72°C for 10 min, step 6: 10°C for 10 min. The mixture used was 2,5 µl of Master Mix (Qiagen Multiplex PCR Kit), 2 µl Primer mix with 1 µl forward and 1 µl reverse combined and 0,5 µl DNA (table 1).

Table 1: Volumes of reagents used for primer PCR

Reagent	**Volume**
Master Mix (Qiagen Multiplex PCR Kit)	2,5 µl
Primer Mix (Combinations VF2_t1-VR1_t1)	2 µl
DNA	0,5 µl
Total volume	5 µl

After running the PCR in a T100 Thermal Cycler (Bio-Rad Laboratories, USA), a gel was prepared with 1,5 % agarose, stained with 0,96 µl Ethidium bromide nucleotide dye and

PeqGold 100bp DNA ladder (VWR Chemicals, USA) was added. The primers were checked for quality in the gel iX20 Imager, samples which needed to be repeated were determined.

2.2.3. Multiplex:

33 primers were developed by Dr. Meimberg at the Institute for Integrative Nature Conservation Research at the University of Life Sciences Vienna, based on an *Oreochromis niloticus* individual. These 33 primers were pooled together into four primer mixes, called mtT|1a, mtT|1b, mtT|2 and mtT|3, which were then pooled into 4 384 well plates after being mixed with Master mix.

Table 2.: one plate of four, containing the first 96 samples with four Primer mixes. g: genomic; number after DNA: number of the sample; mtT|1a-mtT|3: name of the primer mix.

	1	**2**	**3**	**4**	**5**
A	gDNA1mtT|1a	gDNA1mtT|1b	gDNA2mtT|1a	gDNA2mtT|1b	…
B	gDNA1mtT|2	gDNA1mtT|3	gDNA2mtT|2	gDNA2mtT|3	…
C	gDNA13mtT|1a	gDNA13mtT|1b	gDNA14mtT|1a	gDNA14mtT|1b	…
D	gDNA13mtT|2	gDNA13mtT|3	gDNA14mtT|2	gDNA14mtT|3	…
E	…	…	…	…	…

The genomic DNA from all samples was added to every primer mix while on an ice box (table 2). A plastic layer was added over the plate to prevent evaporation. After a short centrifugation the plate was placed into a Thermocycler. The PCR cycler program LT-PCR 30 was used to amplify the primers, running at 95°C denaturation temperature. Annealing was happening at 55°C and elongation at 72°C, which was repeated for 30 cycles. The product was quality controlled via gel electrophoresis on a 1,5 % agarose gel, which was stained with 4 µl of Ethidium bromide nucleotide dye and 1,5 µl of DNA primer product. A PeqGold 100 bp DNA ladder (VWR Chemicals, USA) was added as reference. A gel iX20 imager was used to check the gel. All those steps were repeated for the four plates.

2.2.4. Purification

After the annealing of the primers, the PCR plates were purified with magnetic beads to remove excess DNA and loose strands as well as unwanted fragments. The mixture containing

the beads (Beckman Coulter Life Sciences, USA) was added to the prepared PCR plates and mixed with an electric multi-channel pipette up to ten times. The PCR-beads mix was left to incubate at least for five minutes at room temperature. After this step, an inverse magnetic separator was covered with a sample plate and moved around in the mixture for two minutes by hand. Next, the magnetic separator was dipped into two plates containing 80% ethanol for washing. The beads were then air dried for 5 minutes. After drying, everything was placed in a 10 mM elution buffer and the magnetic separator was removed, leaving only the plate to be moved around in the buffer for 2 minutes to transfer the DNA into the buffer. The separator was placed back onto the plate and moved around again for at least 2 minutes and until the solution was clear to separate the beads from the DNA. After this step, the Illumina-Indexing was done.

2.2.5. Illumina library preparation: Index PCR

In the second PCR step, the unique index combinations were added to the DNA fragments. This step is necessary to identify the individual samples after they have been pooled together, so to every sample, a unique combination of indexing primers is added (Illumina library). The volumes used are described in table 3.

Table 3: Reagents used for the Illumina library PCR

Reagent	**Volume**
Master Mix (Qiagen Multiplex PCR Kit)	5 µl
Primer P5, 1 µM	2 µl
Primer P7, 1 µM	2 µl
PCR Product	1 µl
Total volume	10 µl

2.2.6. Primer Mix Schemes

To prevent P5 and P7 primers from mixing before adding to the PCR product, they were pipetted on opposite sides of each well. PCR indexing was done in two runs. The primer combinations are unique, as the primers P5C1-P5C24 are combined with the primers P7CS1-P7CS32 in such a way, that every sample has its own primer combination and can be

unambiguously identified, after the samples have been pooled together into one vial and sent in for Illumina sequencing.

The BioRad T^{TM}Thermal Cycler was again utilised for ligating the index primers in the following steps: step 1: 15 minutes at 95°C denaturation, step 2: 95°C for 30 sec, step 3: 58°C annealing for 1 min, step 4: 72°C elongation for 1 min. 10 repeats for the steps 2 – 4, step 5: 72°C for 5 min.

The mixture used was 5 µl of Master Mix (Qiagen Multiplex PCR Kit), 4 µl Primer mix with 2 µl forward and 2 µl reverse combined and 1 µl DNA

For quality control a 1,8% agarose gel was prepared for gel electrophoresis with 3 µl of Ethidium bromide and loaded with 4 µl loading dye and 1µl PCR product. The samples were then pooled and sent to Illumina with the MiSeq modality.

2.2.7. PCR for Sanger Sequencing

The PCR protocol for Sanger sequencing was the same procedure as described above, both for the reactant volumes as well as the PCR cycle setting. Primer pair used for Sanger sequencing was VF2_t1, VR1_t1 (table 4, Ward et al. 2005). Quality of PCR product was tested on a 1,8% agarose gel with the same additives as mentioned above.

Table 4: Cytochrome Oxidase Subunit 1 Primers used for Sanger Sequencing with respective Nucleotide sequence

Primer name	Nucleotide Sequence
VF2_t1	CAACCAACCACAAAGACATTGGCAC
VR1_t1	TAGACTTCTGGGTGGCCAAAGAATCA

2.3. Data processing

2.3.1. Data preparation

The data from the Illumina MiSeq is received in the form of Fastq files. The sequences are presented in four lines, first a header containing the name of the sequence. The second line is the sequence. The third line can be used for comments on the sequence and the fourth line provides quality data for each read. Two reads are done per sample: one in each direction

which are then available as two separate sequences. R1 and R2 are read one in forward direction and read two in reverse direction respectively. In several steps, data preparation and analysis were done with the help of numerous scripts and software programs.

To check for low quality sequences, Fastqc (Andrews, 2010) was used. These sequences were excluded. Cutadapt (Martin, 2011) was used to trim parts of sequences which were of low quality. The forward and reverse direction sequences were merged to form a single coherent sequence per sample with PEAR (Zhang, et al. 2014). Centroid clustering was done with Vsearch. Potential Operating Taxonomic Units (OTUs) with low frequencies (below 10) were excluded, as they are probable PCR or sequencing errors.

2.3.2. Sanger data processing

The raw data was analysed with Genious 9.2.3 and sequences were allocated to proper read direction and amplicons were merged with the De Novo Assemble tool. Sequence quality was checked and the peaks in the electropherogram were inspected. Amplicons were trimmed after merging to 650 bp by cutting the primer sequences. Ambiguous bases were marked with 'N'.

2.3.3. Illumina data processing

The data from Illumina sequencing was processed in multiple steps, using the protocol for non-overlapping paired end reads: first, the sequences with good quality were merged using Vsearch and consensus sequences were made with a minimum length of 300 bp. Remaining unmerged sequences were defined as reverse strands and merged with the command 'fastx revcomp'. The gaps between the then merged sequences were filled with ambiguous nucleotides, following the IUPAC code for consensus sequences.

Using a custom python script, provided by Dr. Manuel Curto, formerly of the Institute for Integrative Nature Conservation, BOKU Vienna, the primer sequences were verified and removed from the consensus sequences. Creating clusters of similar sequences with Vsearch, the clustering threshold was considered to be an additional quality criterion for the sequences. The fasta files resulting from this step were summarized and used for the next working steps.

2.4. Data analysis

To analyse the data and visualize differences in the sequences and haplotypes, the sequences were concatenated into files on a genus level and run through various programs:

The sequences were tested by BLAST search on the GenBank databases of NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and BOLD (https://boldsystems.org/index.php). Next, minimum spanning networks were constructed using PopArt.

2.4.1 Construction of PopArt minimum spanning networks

For PopArt (Bandelt et al. 1999) to be able to read the sequences of haplotypes, they had to be converted into Nexus files which was done at https://www.hiv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html and provided with additional information like a characteristics matrix for differentiating haplotypes from different catchments. Sequences that had longer sections of ambiguous nucleotides had to be deleted due to the software excluding all other sequences in this region. Likewise, regions of sequences that shared ambiguous nucleotides had to be deleted as well. The remaining regions were then used to construct minimum spanning networks to visualize haplotype differences.

2.4.2 Construction of sequences with 33 Primers and VF2 t1, VR1 t1 Sequences and Blasting

Haplotypes from the Illumina sequencing was fused with haplotypes from the Sanger sequencing in BioEdit (Hall, 1999) for Mac by hand. The COI region in the mitochondrial genome, which the Sanger Sequences consist of were cut into the Illumina sequences where data was missing or of insufficient quality. In this step, sequences of poor quality were excluded as well.

A database of barcode index numbers (BINs) of chordata was downloaded from ncbi.nlm.nih.gov. The dataset of haplotypes was blasted against this dataset for confirmation of species identification.

2.4.3 Creating a Neighbour Joining Tree

As a next step, the COI sanger sequences with the primers VF2_t1, VR1_t1 were run through Geneious 8.1.9 (Kearse et al. 2012) software by Dr. Lara Baptista to create a phylogenetic tree in the form of a neighbour joining tree. This was done in order to calculate and visualize relationships between the species.

2.4.4 Finding BINs for sequences

The sequences from Sanger sequencing that were used to form a neighbour joining tree were individually run through the BOLD (Barcode of Life Data System, boldsystems.org) library to identify the according BINs (Barcode Identification Number) and to check for inconsistencies in identification and sequence-usability.

3. Results

3.1 Species Composition

3.1.1 Species summary

411 Fish were caught in total. 14 families containing 27 genera were caught, of which 33 species could be determined morphologically. The most diverse family was the Alestidae with 8 different species, second most diverse was the Cichlidae with 6 different species. Total number of catches for the Comoé catchment was 112, for the Mouhoun catchment it was 299. In the Comoé catchment, 12 different species were caught, while in the Mouhoun the number of species was 29, based on morphological determination. Some individuals could only be determined to genus level and were not included in the species count. The family that is represented in the highest abundance are the Cichlidae with 177 individuals, second most abundant the Alestidae with 116 individuals. Rare species like *Petrocephalus soudanensis, Synodontis membranaceus, Auchenoglanis occidentalis, Ctenopoma kingsleyae, Bagrus bajad and Heterotis niloticus* were only caught once (table 5). In regard to the genus *Synodontis*, morphological determination suggests 4 caught species. Molecular methods however suggest the presence of a cryptic species within *Synodontis schall* which will be discussed in more detail later.

Table 5: Summary of morphologically identified species in two catchments in three sampling areas.

Family (N=14)	**Genus (N=27)**	**Species (N=33)**	**Sampling Catchment (N=2) Sampling Area (N=3)**			**N of fish per sp.**
			Comoé	**Mouhoun**		
			Banfora	Bobo	Boura	
Alestidae	Brycinus	*B. macrolepidotus*		1	2	3
		B. nurse		17	7	24
	Micralestes	*M. occidentalis*		7		7
		M. sp.			2	2
	Rhabdalestes	*R. septentrionalis*		12	1	13
	Enteromius (former: Barbus)	*E. macrops*	14	17	10	41
		E. sublineatus		14		14
	Labeo	*L. senegalensis*	1		5	6
		L. sp.		1		1
	Raiamas	*R. senegalensis*		3		3
Alestiadae sp.				2		2
Mormyridae	Hippopotamyrus	*H. pictus*			2	2
	Marcusenius	*M. senegalensis*	3			3
	Petrocephalus	*P. soudanensis*			1	1
Mochokidae	Synodontis	*S. membranaceus*			1	1

		S. nigrita		12	18	30
		S. punctifer		2		2
		S. schall			12	12
Cichlidae	Chromidotilapia	*C. guntheri*	6			6
	Coptodon	*C. zillii*	18	13	9	40
	Hemichromis	*H. bimaculatus*	7	9	11	27
		H. fasciatus	17	7	10	34
	Oreochromis	*O. niloticus*	16	5	14	35
	Sarotherodon	*S. galilaeus galilaeus*	11	11	13	35
Claroteidae	Auchenoglanis	*A. occidentalis*			1	1
	Chrysichthys	*C. auratus*	10	1	2	13
		C. sp	1			1
Distichodontidae	Distichodus	*D. rostratus*			1	1
	Paradistichodus	*P. dimidiatus*		4	9	13
Clariidae	Clarias	*C. anguillaris*		8	3	11
		C. sp.	1	1		2
Schilbidae	Schilbe	*S. intermedius*		12		12
Anabantidae	Ctenopoma	*C. kingsleyae*		1		1
Bagridae	Bagrus	*B. bajad*			1	1
Polypteridae	Polypterus	*P. endlicherii*			2	2
Arapaimidae	Heterotis	*H. niloticus*	1			1
Channidae	Parachanna	*P. obscura*		2		2
Latidae	Lates	*L. niloticus*	6			6
		Total N Samples				411
		Species per area	12	20	22	
		Total N Species	33			
		Total N Genus	27			
		Total N Families	14			

3.1.2 Occurrence of species

Enteromius macrops occurred at 9 sampling sites. *Sarotherodon galilaeus galilaes* occurred at 7 sites. *Oreochromis niloticus* and *Hemichromis fasciatus* occurred at 6 sampling sites. *Hemichromis bimaculatus* and *Coptodon zillii* occurred at 5 sites.

3.1.3 Species composition Comoé

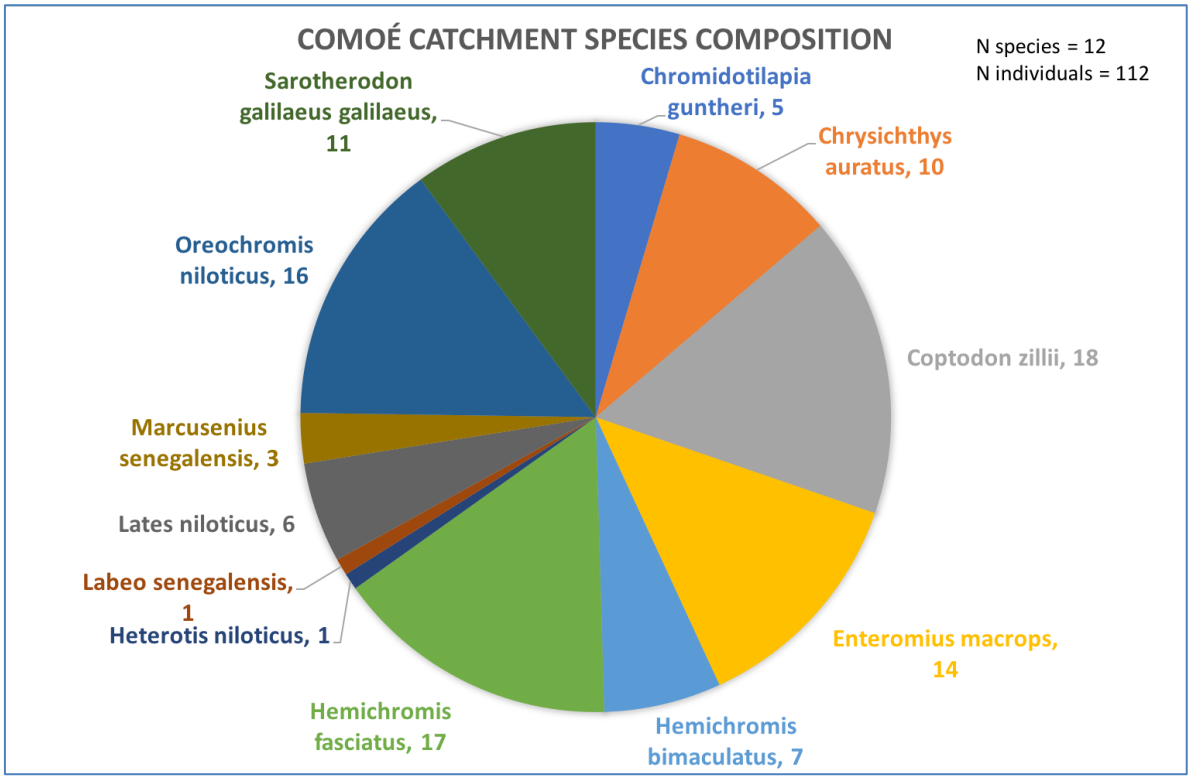

Figure 8: Pie chart species composition for Comoé catchment with 112 caught individuals, in 12 morphologically determined species. Names of the species are shown next to the number of caught individuals

For the Comoé catchment, the species with the highest abundance was *C. zillii* with 18 caught individuals, followed by *H. fasciatus* with 17 and *O. niloticus* with 16 individuals. *E. macrops* was caught 14 times, *S. galilaeus* 11 and *C. auratus* 10 times. *H. bimaculatus, L. niloticus, C. guntheri, M. senegalensis, L. senegalensis* and *H. niloticus* were caught less than 10 times with the latter two only once. The number of caught individuals was 112 (figure 8)

3.1.4 Species composition Mouhoun

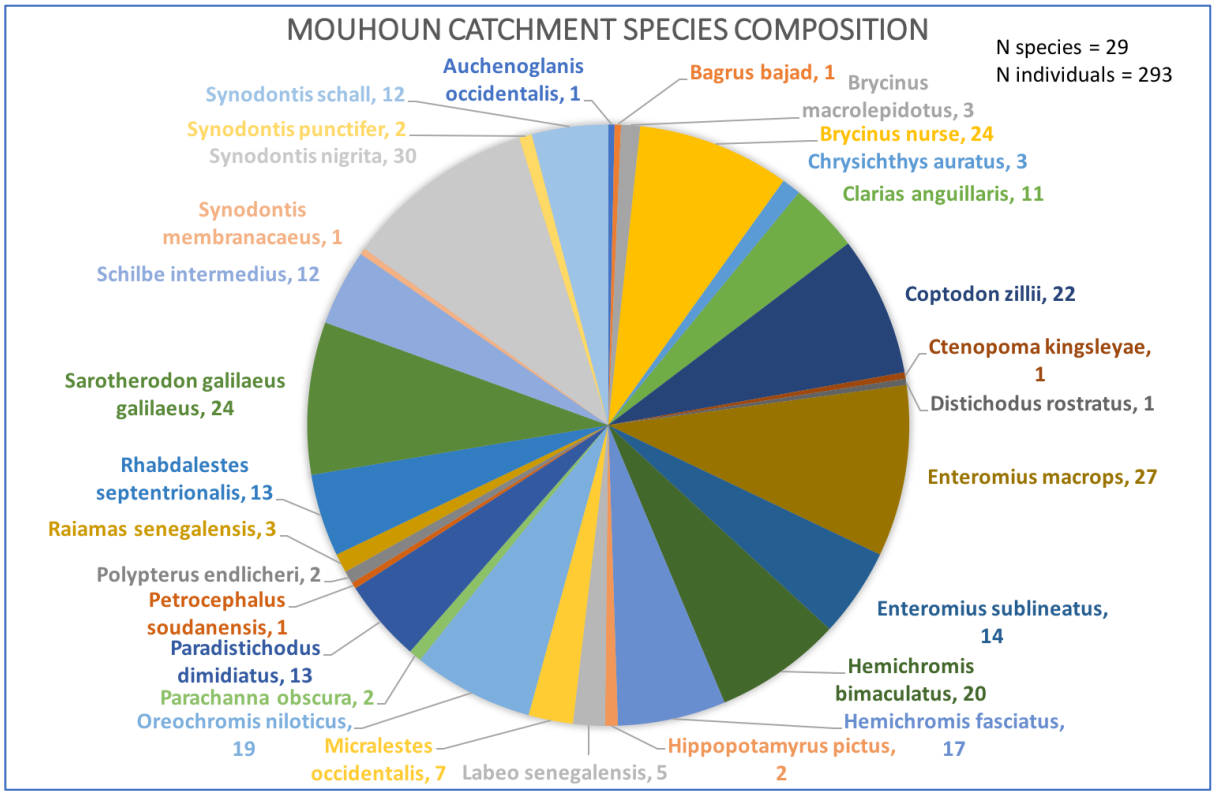

Figure 9: Pie chart species composition for Mouhoun catchment with 293 caught individuals, in 29 morphologically determined species. Names of the species are shown along the number of caught individuals.

The Mouhoun species composition is composed of 29 morphologically identified species. The species with the most abundance according to catches was *Synodontis nigrita* with 30 individuals, second most abundant species was *E. macrops* with 27 individuals. *S. galilaeus* and *B. nurse* were both caught 24 times. *H. bimaculatus* was caught 20 times, *O. niloticus* 19 times, *H. fasciatus* 17 times, *E. sublineatus* 14 times, *R. septentrionalis and P. dimidiatus* 13 times, *S. intermedius and S. schall* 12 times and *C. anguillaris* 11 times. *M. occidentalis* is present with 7 individuals, *L. senegalensis with* 5 individuals. *R. senegalensis, C. auratus and B. macrolepidotus are present* 3 times. *H. pictus, S. punctifer, P. endlicheri and P. obscura* were caught twice, *D. rostratus, C. kingsleyae, B. bajad, A. occidentalis, S. membranacaeus and P. soudanensis* once (figure 9).

3.2. NGS and SANGER sequence data

3.2.1. Haplotype Neighbour joining tree

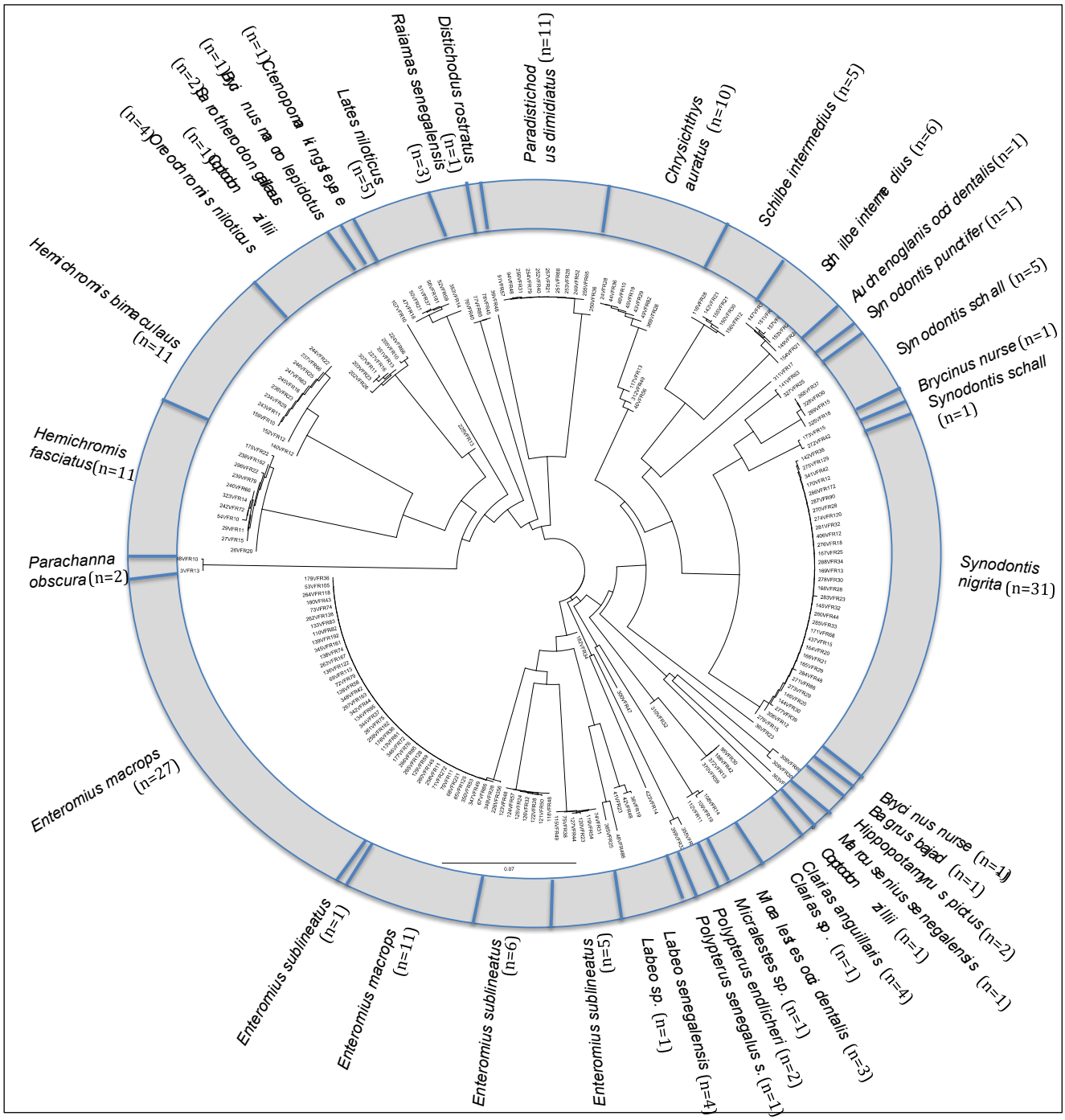

Figure 10: Neighbor joining tree from COI sequences of 194 samples collected in the Mouhoun and Comoé catchments. Not included are the species: Rhabdalestes septentrionalis, Chromidotilapia guntheri, Petrocephalus soudanensis, Alestidae sp., Heterotis niloticus.

For Sanger sequencing data, the Haplotypes for most species are forming one terminal branch indicating clear species assignment (figure 10).

Cichlidae

Different species comprising very similar haplotypes were observed in the Cichlids. The Cichlid family appears to be presented by three distinct clades, two include the *Hemichromis* genus with the two species *H. fasciatus* and *H*. *bimaculatus* each on one clade but comprising several closely related haplotypes. The third branch includes three species: *Oreochromis niloticus, Coptodon zillii, and Sarotherodon galilaeus,* (subfamily Pseudocrenilabrinae) which share one clade of haplotypes, indicating low genetic structure between the species concerning mitochondrial DNA. This indicates the possibility of occasional hybridization between the three species that were formerly treated as congeneric (Tilapia; e.g. Tibihika et al., 2019).

Anabantidae & Channidae

Ctenopoma kingsleyae and *Parachanna obscura* of the order Anabantiformes are present in distinct branches neighbouring the Cichlids. One *C. zillii* individual is positioned deviating from these clades. Because it has an own haplotype differentiated by high distance from the next neighbour the individual was probably misidentified.

Cyprinidae

Shared haplotypes between close related species were identified also for the species *E. macrops* and *E. sublineatus* with one individual. This can be due to misidentification or occasional genetic exchange. Four species showing a clear differentiation in haplotypes exceeding the distance normally associated with species differences. *E. sublineatus* samples form a group with two clearly differentiated haplotypes of about 12% sequence divergence. *Labeo senegalensis* shoulders the *Enteromius* haplotypes with two branches on its own, indicating probably misidentification since its main distinction from *L. coubie* is variation in colour. *Raiamas senegalensis* is present with three individuals but located between Latidae and Distichodintidae.

Distichodontidae

Distichodus rostratus with a distinct haplotype is present with one individual. *Paradistichodus dimidiatus* is neighbouring the haplotype with eleven sequences.

Schilbeidae

As it is with Enteromius, the same is the case with *Schilbe intermedius*: in both cases the two groups occur within one catchment indicating the existence of additional sympatric species that have not been recognized morphologically.

Claroteidae

Chrysichthys auratus forms two groups of haplotypes. Here the groups are correlated to the two catchments and could indicate differentiation between them.

Bagridae

Bagrus bajad is the only member from the Bagridae family present, with one sample from the Mouhoun catchment.

Clariidae

Clarias anguillaris is present with five individuals, one of which could not be identified morphologically with confidence. Hence it was labelled as *Clarias sp.* This was also the only *Clarias* from the Comoé catchment.

Mochokidae

Synodontis schall is divided into two groups. The Synodontidae appear therefore as four groups while only three are morphologically identified. Hereby, *Synodontis punctifer* and *Synodontis nigrita* present one group respectively.

Mormyriade

Hippopotamyrus pictus and *Marcusenius senegalensis* are Mormyriade present with three samples and a *Coptodon zillii* whose haplotype is equidistant from the previous two. This is likely due to an error in the sequence, since they are highly dissimilar in morphology and not likely to be mistaken for each other.

Alestidae

Micralestes occidentalis, a member of the Alestidae family is present with three sequences. One Micralestes individual could not be determined to species level morphologically.

The positions of another Alestidae, *Brycinus nurse* in these groups are probably erroneous, which also applies to *Brycinus macrolepidotus.*

Polypteridae

Polypterus senegalus senegalus and *Polypterus endlicheri* are Polypteridae whose haplotypes are neighbouring and branching off from each other.

The haplotype differentiation between catchments within a species group has to be analysed further.

3.2.2. Haplotype Minimum Spanning Networks

Synodontis schall

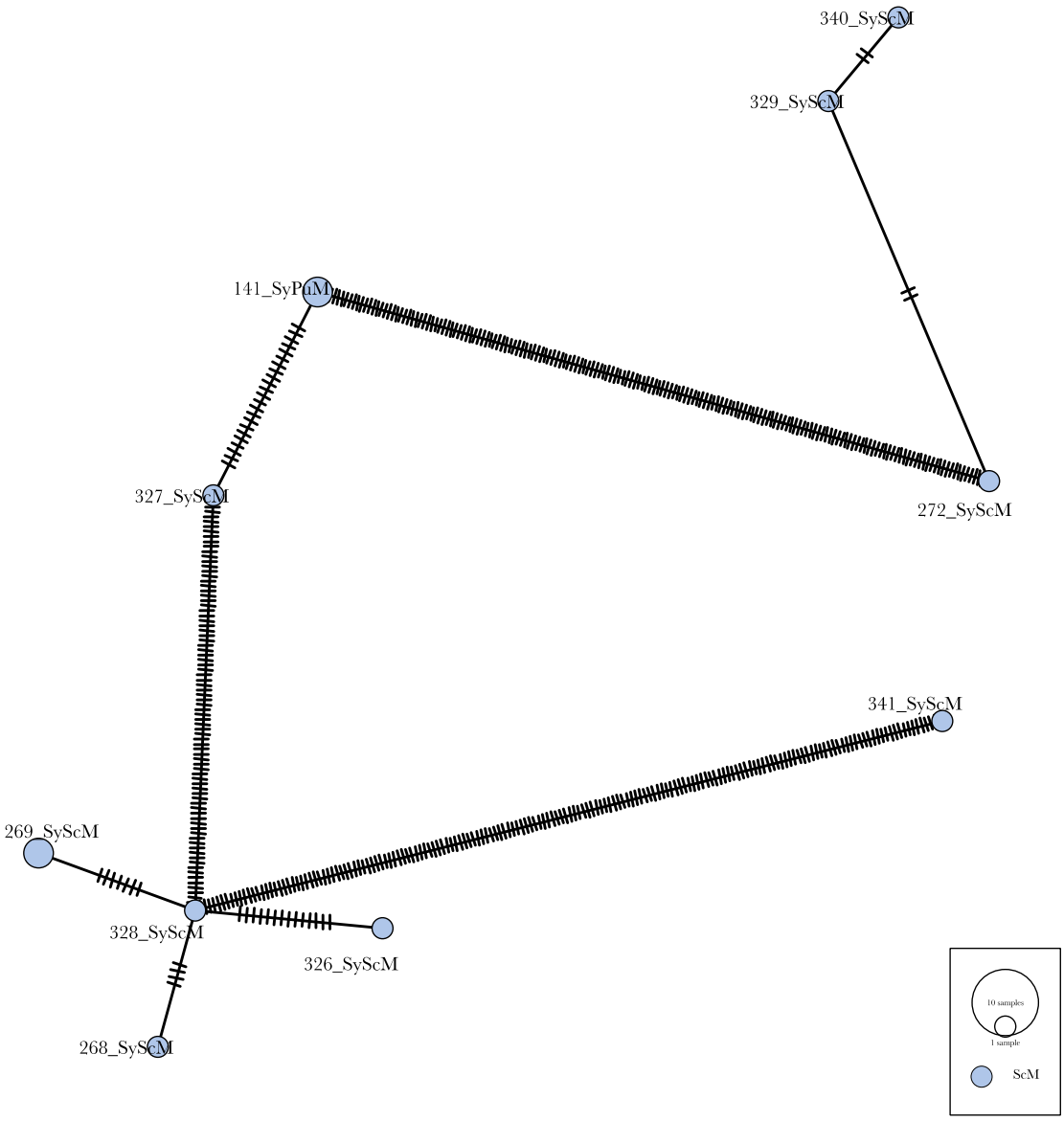

Figure 13: Minimum spanning network for Synodontis schall from Mouhoun catchment, with NGS sequence data. Number of samples: 14

The 12 sequences of *Synodontis schall* are comprised of 10 different haplotypes, some of which are only separated by 2-14 differences, others considerably more. Sample number 141, which was morphologically identified as *Synodontis punctifer* has the same sequence as

sample 330, which, like the rest, was identified as *S. schall.* Sample 327 is separated by a few differences in nucleotide makeup from sample 141 and 330 in NGS sequence data.

Chrysichthys auratus

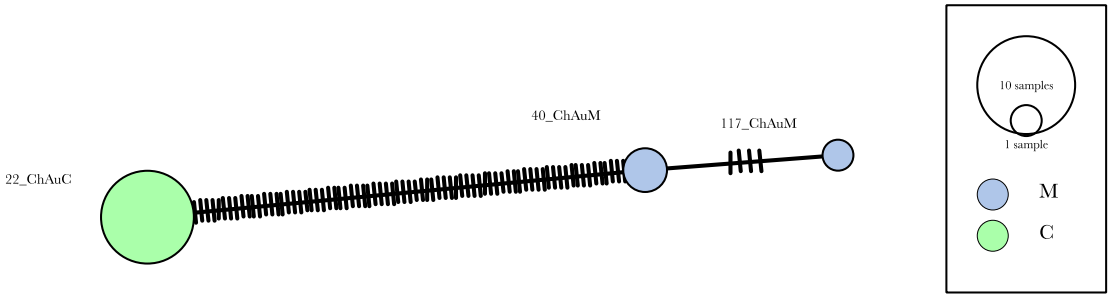

Figure 11.: Minimum spanning network for Chrysichthys auratus. Green circle represents samples from Comoé catchment, blue for Mouhoun catchment. Width of circles is representative for the number of samples grouped to one haplotype, which are 9, 2 and 1 from left to right respectively. Sequences used were NGS and Sanger sequences combined. Number of samples: 10.

Chrysichthys auratus shows a considerable amount of base changes between one haplotype group from Comoé catchment (samples 22, 23, 24, 43, 44, 45, 46, 49, 369) and another haplotype group from Mouhoun catchment, which is in itself separated into two small groups by four base changes (samples 40, 312 and sample 117) (figure 11). This shows the same differentiation, namely sample 24, 43, 44, 45, 46, 49, 369, and on the other hand sample 40 and 117, as in the Neighbour joining tree (figure 10).

Schilbe intermedius

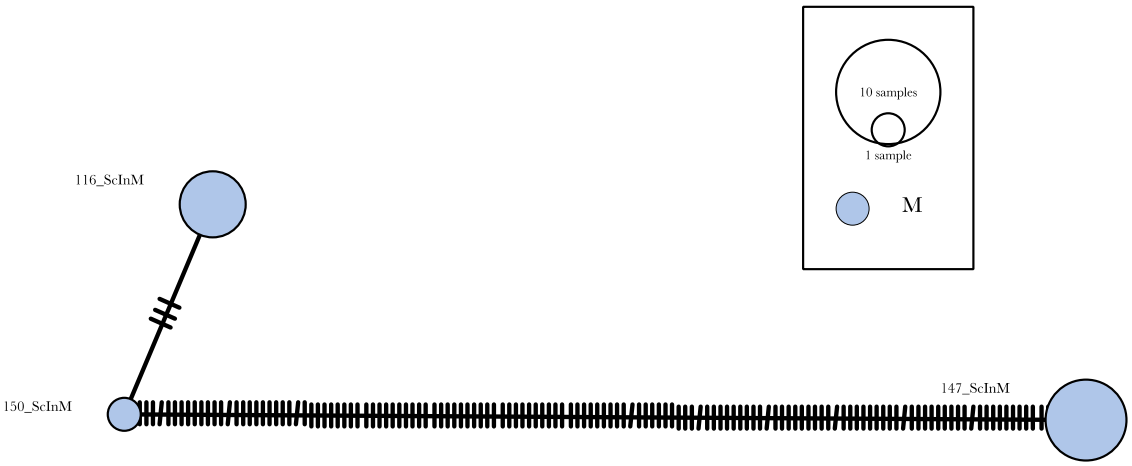

Figure 12: Minimum spanning network for Schilbe intermedius with NGS sequence data. All samples from Mouhoun catchment. Width of circles is representative for the number of samples grouped to one haplotype. Number of samples: 12

A minimum spanning network for *Schilbe intermedius* shows a distinction between two groups of *S. intermedius* samples. One group of haplotypes is separated by three mutations, consisting of sample 116, 143, 150, 155 and 156. The second group consists of samples 147, 149, 151, 153, 154 and 157 (figure 12).

The Neighbour Joining Tree shows the same differentiation in regard to haplotypes for the COI region (figure 10). As all the samples were caught in the Mouhoun catchment, a comparison between catchments for this genus was not possible.

3.2.3. BIN search and haplotype checking

Of the 29 species identified morphologically which had COI sequences of Sanger Sequencing, 18 species could be identified correctly by checking with BOLD (Barcode of Life Data Systems). Of those 18 species, 14 identifications could be made with enough accuracy to find a BIN (Barcode Index Number) corresponding to the identification on both morphological and molecular level. Two BINs were found that did not match with morphological identification. BOLD:ACR5963, *Enteromius macrops* was found for sample nr. 138, which was morphologically identified as *Hemichromis bimaculatus* and was assigned haplotype nr. 30. BOLD:ACH7500, *Synodontis violaceus* was found for sample nr. 141, Synodontis punctifer as well as sample nr. 327*, Synodontis schall,* both of which were assigned haplotype nr. 63. Furthermore, some samples that were morphologically different showed up in the same haplotype group: sample nr. 341 *Synodontis schall* together with *Synodontis nigrita* in haplotype nr. 62 or sample nr. 129 which was *Enteromius sublineatus* and got lumped together with *Enteromius macrops* in haplotype group nr. 24 (Table 6). For a considerable number of haplotype sequences, the TOP HIT identification was coherent with the morphological identification, but a close enough matching BIN (within 3%) could not be found.

Table 6: Haplotypes with according morphological identification, sample number as representative for arbitrary sample per haplotype. Catchment: M=Mouhoun, C=Comoé. Top Hit: Most likely species. BIN: when present, species identified with 3% confidence. Top %, Low %: Similarity scores in percentage of top 100 matches of sequence.

Haplotype NR	N=	Morphological species	TOP HIT	BIN	Top %	Low %
1	1	Brycinus nurse	Synodontis ocellifer		99,8	93,86
2	2	Brycinus nurse	Synodontis nigrita		99,49	96,72
3	2	Brycinus nurse	Sarotherodon galilaeus		100	99,40
4	3	Brycinus nurse	Oreochromis sp.		100	100
5	1	Chromidotilapia guntheri	Chromidotilapia guntheri	BOLD:AAL6045	99,21	86,69
6	4	Chromidotilapia guntheri	Chromidotilapia guntheri	BOLD:AAL6045	99,37	86,69
7	3	Chrysichthys auratus	Chrysichthys sp.		98,88	95,38
8	7	Chrysichthys auratus	No match		NA	NA

9	1	Chrysichthys auratus	Sarotherodon galilaeus		98,88	95,52
10	1	Clarias anguillaris	Oreochromis niloticus		100	99,80
10	1	Clarias sp.	Oreochromis sp.		100	98,22
11	5	Clarias anguillaris	Clarias gariepinus		100	99,60
12	7	Coptodon zillii	Coptodon zillii (Tilapia zillii)		100	99,53
13	1	Coptodon zillii	Coptodon zillii		99,80	99,35
14	3	Coptodon zillii	Coptodon zillii		99,69	99,22
15	3	Coptodon zillii	Coptodon zillii		99,80	99,8
16	1	Coptodon zillii	Oreochromis sp.		100	100
17	1	Coptodon zillii	Oreochromis niloticus		100	99,80
18	1	Coptodon zillii	Oreochromis niloticus		100	99,40
19	2	Coptodon zillii	Sarotherodon galilaeus		100	99,40
20	4	Coptodon zillii	Sarotherodon glailaeus		99,60	85,34
21	1	Coptodon zillii	Hemichromis saharae		NA	NA
22	7	Enteromius sublineatus	NA		NA	NA
23	5	Enteromius sublineatus	NA		98,46	91,89
24	38	Enteromius macrops	Enteromius macrops	BOLD:ACR5963	97,96	91,65
24	1	Enteromius sublineatus	Enteromius macrops	BOLD:ACR5963	99,60	85,34
25	1	Enteromius macrops	Enteromius macrops	BOLD:ACR5963	99,80	85,14
26	3	Hemichromis bimaculatus	Hemichromis saharae		100	92,99
27	20	Hemichromis bimaculatus	Hemichromis saharae		100	100
28	9	Hemichromis fasciatus	Hemichromis bimaculatus		98,29	91,84
29	3	Hemichromis fasciatus	Oreochromis sp.		100	99,40
29	1	Hemichromis fasciatus	Sarotherodon galilaeus		100	84,55
30	1	Hemichromis bimaculatus	Enteromius macrops	BOLD:ACR5963	NA	NA
31	1	Hemichromis bimaculatus	Sarotherodon galilaeus		99,40	84,81

32	1	Hemichromis fasciatus	Hemichromis bimaculatus		99,40	84,94
33	1	Hemichromis fasciatus	NA		NA	NA
34	1	Hemichromis bimaculatus	Hemichromis saharae		NA	NA
35	1	Hemichromis bimaculatus	Hemichromis saharae		100	93,87
36	1	Hippopotamyrus pictus	NA		100	92,27
37	1	Hippopotamyrus pictus	NA		99,80	92,35
38	3	Labeo senegalensis	Labeo vulgaris		100	95,47
39	1	Labeo senegalensis	Labeo coubie		NA	NA
40	1	Labeo sp.	Labeo coubie	BOLD:AAL6302	100	100
41	5	Lates niloticus	Lates niloticus	BOLD:AAA2960	NA	NA
42	1	Micralestes occidentalis	NA	NA	99,2	87,62
43	1	Micralestes occidentalis	Oreochromis sp.		100	99,60
44	1	Micralestes occidentalis	NA	NA	100	99,80
45	6	Micralestes occidentalis	Micralestes humilis		99,80	99,80
46	1	Oreochromis niloticus	Coptodon zillii (Tilapia zillii)		100	99,80
47	1	Oreochromis niloticus	Oreochromis niloticus		100	87,97
48	18	Oreochromis niloticus	Oreochromis sp.		100	87,43
49	9	Oreochromis niloticus	Oreochromis niloticus		100	87,28
50	11	Paradistichodus dimidiatus	Paradistichodus dimidiatus	BOLD:ACS6030	100	89,70
51	1	Polypterus endlicheri	Polypterus endlicheri	BOLD:AAY3429	100	88,33
52	1	Polypterus endlicheri	Polypterus endlicheri	BOLD:AAY3429	99,80	99,20
53	1	Polypterus senegalus	Polypterus senegalus	BOLD:AAH7765	100	99,40
54	3	Raiamas senegalensis	Raiamas senegalensis	BOLD:ACR9171	100	100
55	1	Sarotherodon galilaeus	Sarotherodon galilaeus		100	92,21

56	28	Sarotherodon galilaeus	Sarotherodon galilaeus		99,80	92,09
57	1	Sarotherodon galilaeus	Oreochromis sp.		99,80	92,26
58	5	Schilbe intermedius	Schilbe intermedius		99,80	92,26
59	6	Schilbe intermedius	Schilbe intermedius	BOLD:AAL5704	100	92,32
60	1	Synodontis nigrita	Synodontis nigrita		100	95,32
61	1	Synodontis nigrita	Synodontis nigrita		100	95,15
62	29	Synodontis nigrita	Synodontis nigrita		100	94,07
62	1	Synodontis schall	Synodontis nigrita		97,27	88,17
63	1	Synodontis puntifer	Synodontis violaceus	BOLD:ACH7500	99,79	88,18
63	1	Synodontis schall	Synodontis violaceus	BOLD:ACH7500	100	92,16
64	4	Synodontis schall	Synodontis schall		100	84,13
65	1	Synodontis schall	Synodontis ocellifer		100	91,48
66	1	Auchenoglanis occidentalis	Auchenoglanis occidentalis	BOLD:AAL5844	99,8	93,86
67	1	Bagrus bajad	Bagrus bajad	BOLD:AAL7132	99,49	96,72
68	1	Marcusenius senegalensis	Marcusenius senegalensis	BOLD:AAL6601	100	99,40
69	1	Parachanna obscura	Parachanna obscura		100	100
70	1	Distichodus rostratus	Distichodus rostratus	BOLD:AAL6019	99,21	86,69

4. Discussion

Overall 411 samples were caught in 13 sites. 33 species could be identified morphologically with 4 instances where additional species are possible. These seem to be formerly unrecognised and demand closer morphological and molecular investigation. The 33 species are made up of 27 genera and 14 families. The Comoé catchment held 12 species with 112 of the total catches while in the Mouhoun catchment, 29 species were caught with 293 individuals.

The higher number of species in one catchment is therefore due to the higher number of overall samples caught there. The unevenness of samples was partly due to difficulties in the sampling itself, where access to sampling sites was often the decisive factor of whether or not sampling was done. Changes to the field work schedule was another reason.

Research question number two, comparing genetic distances of selected species between river catchments, could not be answered confidently due to a lack of samples and will be discussed in more detail further on.

Inconsistencies within the Oreochromis, Coptodon and Sarotherodon genera with regard to their shared haplotypes are within the expected range. Occasional hybridization between the species is suspected and made morphological differentiation difficult in some instances (Tibihika et al. 2019).

Haplotype sharing with *Enteromius sublineatus* and *Enteromius macrops* are potentially due to misidentification, as the species only grows to a length of about 7 cm (West Africa Fish identification key). Another source for potential errors in the haplotype differentiation is the possibility of genetic exchange between the closely related species. With 12% sequence divergence between the haplotypes, *Enteromius sublineatus* seems to harbour a cryptic species which is not recognised morphologically. Molecular identification was problematic as well, since the BOLD library does not hold records for this species, which will be discussed later on. Samples need to be checked further and the specimens have to be revisited.

For *Schilbe intermedius* the haplotypes indicate the existence of two species that have been recognised as one morphologically. A minimum spanning network also shows a

corresponding grouping of two haplotypes, separated by numerous base pairs (figure 14). The species was only caught in the Mouhoun catchment. Calculating the identity/similarity for two sequences from the respective haplotypes results in an identity of 0,8922801. This equates to a sequence difference of 10,77%. The number of base differences between the haplotype groups of *Schilbe intermedius* is visualized in figure 14.

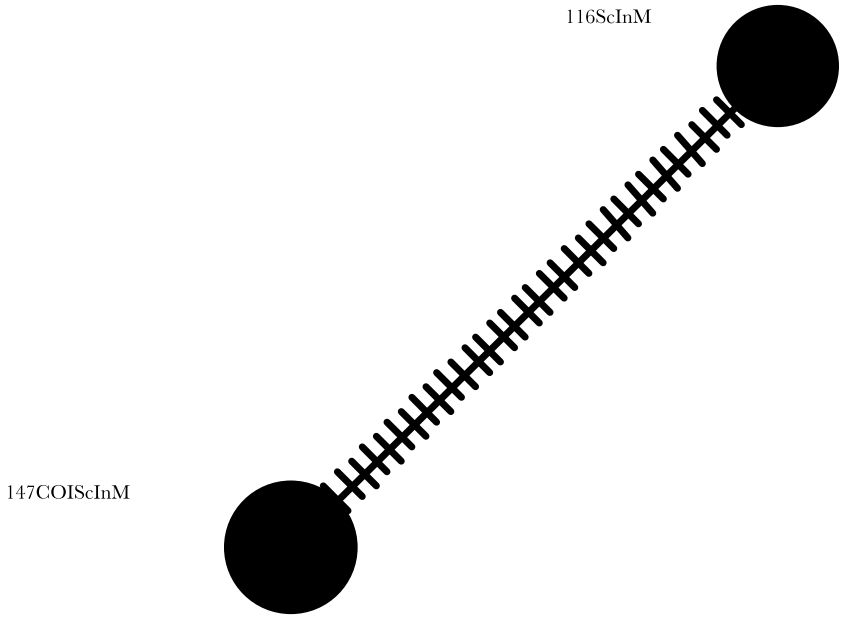

Figure 14: Minimum spanning network, illustrating the difference of haplotypes of two groups of Schilbe intermedius samples. Grouped are: Sample Nr: 147, 149, 151, 153, 154, 157 and samples nr.: 116, 143, 150, 155, 156 corresponding to the neighbour joining tree haplotype branches.

Chrysichthys auratus is seemingly made up of two haplotypes, with the samples that form a distinct branch being from a different catchment. This can indicate allopatric species due to the different catchments but can also be due to a cryptic species being treated as one morphologically. When comparing COI sequences identities/differences from two terminal branches, namely the sequence from sample 117 and from sample 369, the identities lie at 0,8635548, which amounts to a difference in the sequences of 13,46 %. Claroteidae like *C. auratus* lack scales which makes morphological differentiation sometimes difficult. Furthermore, for other species, haplotype and catchment differentiation do not correlate so an isolated instance should not be taken as anything other than coincidence. Samples 22, 23

and 312 were not included in the Neighbour joining tree due to missing sequence data in the necessary region

The Synodontis genus is the most species rich of the Mochokidae and accounts for about one quarter of all African siluriform species. Given this fact, it is surprising that there is no molecular phylogeny available for this group. Some former attempts to shed light on this genus used the mitochondrial cytochrome *b* gene and found relationships between East African clades and South-African samples (Bruwer et al. 2000). No West African or Central African clades have been studied yet in this manner, so the possibility of placing them into a larger phylogenetic and phylogeographic context is amiss (Koblmüller et al. 2006). To help alleviate some confusion, special emphasis was placed on the genus Synodontis in this study.

Synodontis schall being represented as two groups in the NJ tree can be indicative of presence of cryptic species. The Minimum Spanning Network (PopArt, 1999) for the genus Synodontis shows a high number of base changes for sample 272COISyScM with respect to sample 268COISyScM. Samples 272 and 268 were morphologically identified as *Synodontis schall*. Sample 272 shows a distinct haplotype in the Neighbour joining tree as well. A sample of *Brycinus nurse, nr. 173* is also associated with the second branch of *S. schall*. It's sequence however is identical with *S. schall,* indicating the possibility of contamination or mix-up in labelling. Since *S. schall* is easily confused with other species from the Synodontis family, especially *S. gambiensis*, from which it only deviates in colour, it is reasonable to assume inconsistencies in the taxonomy of this family. When checking both sequences with BOLD, *Synodontis ocellifer* was the result, indicating a possible misidentification for sample nr. 272, *S. schall.*

When checking for identities/differences in both *S. schall* sequences from their terminal branches, a difference of 0,9048474 is detected, which equates to a difference of 9,5 % in the nucleotide sequence. Sample 141 was identified as *Synodontis punctifer.* Sample 327 was identified as *S. schall.* However, both are forming one terminal branch in the NJ tree, indicating clear species assignment. This may indicate a misidentification for sample nr. 327, *S. schall,* as it also neighbours the haplotype group sample nr. 141, *S. punctifer* and nr. 330, *S. schall*, in the minimum spanning network for NGS data (figure 13). If the samples are checked by a specialist, sample nr. 330 *S. schall,* may turn out to be *S. punctifer* as well.

Catchment haplotype differentiation could not be observed as the samples were not sufficiently distributed over the two catchments and individuals were not caught at sufficient abundance to make a meaningful observation. To catch enough individuals of a given species, more time and resources would have been necessary, which, at the time was not possible, as fieldwork was done under a time constraint. Furthermore, the disproportionate sampling of the Mouhoun catchment contributed to an underrepresentation of samples from the Comoé catchment. The reasons for more sampling in this area are the already mentioned rescheduling and the accessibility of sampling sites.

Brycinus nurse could not be identified correctly in a single instance when checking the sequences with BOLD.

For *Enteromius sublineatus,* there were no matches at all when checked, which may have been due to the fact that *E. sublineatus* does not have a record in the BOLD library. The BIN of *Enteromius macrops* has only a single public record in BOLD from Sierra Leone. *Clarias anguillaris* and *Hippopotamyrus pictus* are missing from the library's records as well. Here, a submission of sequences can be helpful for species identification in the future.

Curiously, Cichlid sequences are underrepresented when sequencing was done with primers VF2_t1 and VR1_t1, suggesting a bad compatibility with Cichlids, or at least the species *Ctenopoma kingsleyae, Coptodon zillii, Oreochromis niloticus* and *Sarotherodon galilaeus galilaeus.* Of 180 Cichlid samples, only 33 sequences could be retrieved with Sanger-sequencing for the barcoding region. Of those 33 samples, 24 were *Hemichromis fasciatus* and *Hemichromis bimaculatus.*

As of now, the presence of 33 morphological identified species is supported, with 4 possible species that are not yet recognised. In continuation of the project, sequences for all samples will be generated to further investigate the occurrence of species in Burkina Faso. Careful re-examination of morphological variation should be done to determine whether or not species status can be attributed to the detected groups with inconsistencies or if they should be treated as Evolutionarily Significant Units (ESUs) (Frazer et al. 2001). In any case,

focusing on the pressure that species diversity and water bodies in general are experiencing, these lineages should be taken into consideration when managing water and fish resources.

Conclusion

As a complementary tool for species identification of the ichthyofauna in the Mouhoun and Comoé basins, barcoding has a certain utility. Regarding uncovering inconsistencies within the taxonomy of certain groups and families, especially the Cyprinidae and Mochokidae, identification keys are needed as well as detailed knowledge on morphological and molecular levels. When implementing these tools in combination and in a structured approach, it may considerably improve accuracy of species identification.

In combination with methods like NJ, molecular tools can be useful to shine a light on taxonomic inconsistencies and uncover areas with need of further investigation like *Enteromius sublineatus, Chrysichthys auratus* and *Schilbe intermedius.* In the course of this study, taxonomic inconsistencies within certain groups were indicated and the utility for molecular identification tools was explored.

5. References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman. D. J., (1997): Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25: 3389-3402

Barcode of Life Database: http://www.boldsystems.org/index.php

Banu, P. K. N., Andrews, S. (2014): Harmony Search PSO Clustering for Tumor and Cancer Gene Expression Dataset. *Int. J. Swarm Intell. Res*., 5, 1-21.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Smith, A. J. (2008): Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), 53–59. doi:10.1038/nature07517

Bandelt, H., Forster, P., Röhl, A., (1999): Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1), 37–48.

Bolger, A. M., Lohse, M., Usadel, B. (2014): Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15), 2114–2120.

Bouda, S., (2002): Généralités sur la pêche au Burkina Faso de la période coloniale à 1997; in Kabre, G., B., and Magnini, S., F.: Recueil des expériences du projet 'Gestion de la Pêche dans le Sud-Ouest' du Burkina Faso. GTZ & MAHRH, 398 p.

Bruwer E.E., van der Bank F. H., Steyn G.J., Wink M., (2000): Phylogenetic relationships between *Synodontis* species: some preliminary results. *Afr J of Aquat Sci*. 25: 89-92.

Decru, E., Moelants, T., De Gelas, K., Vreven, E., Verheyen, E., Snoeks, J., (2016): Taxonomic challenges in freshwater fishes: A mismatch between morphology and DNA barcoding in fish of the north-eastern part of the congo basin. *Molecular Ecology Resources* 16 (1): 342-52.

Fraiture, de, C. M. S., Kouali, G., Sally, N., Kabre, H., (2014): Pirates or pioneers? Unplanned irrigation around small reservoirs in burkina faso. *Agricultural Water Management* 131: 212 - 20.

Froese, R. and D. Pauly. Editors. 2020. FishBase. World Wide Web electronic publication. www.fishbase.org, version (12/2020).

Geneious Prime 2019.2.3.: https://www.geneious.com/

Grill, G., Lehner, B., Lumsdon, A. E., MacDonald, G. K, Zarfl, C., Reidy Liermann, C., (2015): An index-based framework for assessing patterns and trends in river fragmentation and flow regulation by global dams at multiple scales. *Environ. Res*. Lett. 10 015001: 5

Hall, T. A., (1999): BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41: 95-98

Hebert, P. D. N., Cywinska, A., Ball, S., de Ward, J., (2003): Biological identifications through DNA barcodes. *Philos Trans R Soc Lond Biol Sci* 270: 313 - 321.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., (2012): Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. *Bioinformatics*, 28, 1647-1649.

Knebelsberger, T., Dunz, A. R., Neumann, D., Geiger, M. F., (2015): Molecular diversity of Germany's freshwater fishes and lampreys assessed by DNA barcoding. *Molecular Ecology Resources,* 15 (3): 562-72.

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

LPI 2014. Living Planet Index database. 2014. < www.livingplanetindex.org/>. Downloaded on 5 July 2018: 13

Lu, Y., Shen, Y., Warren, W., Walter, R., (2016): Next Generation Sequencing in Aquatic Models, Next Generation Sequencing - Advances, Applications and Challenges, Jerzy K Kulski, *IntechOpen*

Mano, K., (2016): Fish Assemblages and Aquatic Ecological Integrity in Burkina Faso. Doctorate Thesis. University of Natural Resources and Life Sciences, Vienna, Austria: 51

Mardis, E. R., (2013): 'Next-Generation Sequencing Platforms', *The Annual Review of Analytical Chemistry*, 6: 287–303.

Mardis, E. R., (2017): DNA sequencing technologies: 2006–2016. *Nature Protocols*, 12(2): 213–218. doi:10.1038/nprot.2016.182

Martin, M., (2011): Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet Journal,* 17: 10-12. https://doi.org/10.14806/ej.17.1.200

Meulenbroek, P., (2013): Fish Assemblages and Habitat Use in the Upper Nakambe Catchment, Burkina Faso. Master Thesis. University of Natural Resources and Life Sciences, Vienna, Austria: 42

Meier, R., Kwong, S., Vaidya, G., Ng, P. K. L., (2006): DNA Barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, 55: 715–728.

Melcher, A., Peloschek, F., Moog, O., Oueda, A., Meulenbroek, P., Stranzl, S., Komandan, M., Koblinger, T., Trauner, D., Kaboré, I., Ouedraogo, R., Savadogo, M., Sendzimir, J., (2014): Sustainable management of aquatic ecosystems and fish resources in Burkina Faso, West Africa. Book of Abstracts, *Symposium SUSFISH Appear Project,* 56: 9

Melcher, A.H., Ouedraogo, R., Schmutz, S., 2012. Spatial and seasonal fish community patterns in impacted and protected semi-arid rivers of Burkina Faso. *Ecological Engineering,* 48: 117-129.

Nwani, C.D., Becker, S., Braid, H.E., Ude, E.F., Okogwu, O.I., Hanner, R., (2011): DNA

barcoding discriminates freshwater fishes from southeastern Nigeria and provides river system-level phylogeographic resolution within some species, *Mitochondrial DNA,* 22:sup1, 43-51.

Ouedraogo, R., (2010): Fish and fisheries prospective in arid inland waters of Burkina Faso, West Africa. Doctor Thesis. University of Natural Resources and Life Sciences, Vienna, Austria.

Tibihika, P. D., Curto, M., Almayehu, E., Waidbacher, H., Masembe, C., Akoll, P., Meimberg. H., (2019): Molecular genetic diversity and differentiation of Nile tilapia (*Oreochromis niloticus*, L. 1758) in East African natural and stocked populations. *BMC Evolutionary Biology*, 20: 16. doi:10.1186/s12862-020-1583-0

Pereira, L. H. G., Hanner, R., Foresti, F., Oliveira C., (2013): Can DNA barcoding accurately discriminate megadiverse neotropical freshwater fish fauna? *BMC Genetics,* 14 (1): 20-21

Ratnasingham, S., Hebert. P. D. N., (2007): BOLD: The barcode of life data system (www.barcoding.life.org). *Mol. Ecol Notes,* 7: 355 - 364

Sanger, F., Nicklen, S., Coulson, A., (1977): DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* USA, 74(12), 5463–67.

Savadogo, M., Ouédraogo, R., Ouéda, A., Seynou, O., Zerbo, H., Kaboré, C.; Dibloni, H. T., (2014): La liste rouge des poissons du Burkina Faso (The red list of fish species in Burkina Faso). *Book of abstracts*: 17

Schwarz E. R., Mwale M., Hanner R., (2008): A role for barcoding in the study of African fish diversity. *South Afr. J Sci,* 104: 293-298

Venot, J., Torou, B., Daré, W., (2014): Construction Space: Water Agencies and Committees in Burkina Faso. *L'Espace géographique*, volume 43(2), 148-163. https://doi.org/10.3917/eg.432.0148

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., Hebert, P. D. N., (2005): DNA barcoding Australia's fish species. *Philos Trans R Soc Lond*, Series B, 360: 1847–1857. doi:10.1098/rstb.2005.1716

Waters, J. M., Cambray J. A., (1997): Intraspecific phylogeny of the Cape galaxias from South Africa: Evidence from mitochondrial DNA sequences. *J Fish Biol* 50: 1329-1338

Wishart M., Hughes J., Stewart B., Impson D., (2006): An extreme level of intra-specific divergence among Cape Peninsula populations of the Cape galaxias, Galaxias zebratus Castelnau 1861, reveals a possible species complex. *Afr J Aquat Sci* 31: 99-106.

World Bank, Development Research Group. <https://data.worldbank.org/indicator/SI.POV.DDAY?end=2014&locations=BF&start=1994&view=chart>
Obtained on 21st of April 2020.

6. List of figures

7. List of tables

8. Supplements

Appendix 1.: BIN names differentiated into haplotypes and species

Table 7.: Number of samples with BINs per haplotype. Example: *Chromidotilapia guntheri*: number of samples: 5. Sample for haplotype nr. 5 with BIN: 1. Number of samples for haplotype 6 with BIN: 4. BIN found in BOLD for all 5 samples: AAL6045. Blank: no BIN found. NA: no corresponding sequence found.

	N of samples with BINs per haplotype												
Haplotypes	**AAA2960**	**AAH7765**	**AAL5704**	**AAL6045**	**AAL6302**	**AAY3429**	**ACH7500**	**ACR5963**	**ACR9171**	**ACS6030**	**NA**	**(blank)**	**Grand Total**
Brycinus nurse												**8**	**8**
Chromidotilapia guntheri				**5**									**5**
5				1									1
6				4									4
Chrysichthys auratus												**11**	**11**
Clarias anguillaris												**5**	**5**
Clarias sp.												**2**	**2**
Coptodon zillii												**24**	**24**
Enteromius macrops								**39**					**39**
24								38					38
25								1					1
Enteromius sublineatus								**1**				**12**	**13**
22												7	7
23												5	5
24								1					1
Hemichromis bimaculatus								**1**				**26**	**27**
26												3	3
27												20	20
30								1					1
31												1	1
34												1	1
35												1	1
Hemichromis fasciatus												**15**	**15**

Hippopotamyrus pictus												**2**	**2**
36												1	1
37												1	1
Labeo senegalensis												**4**	**4**
Labeo sp.					**1**								**1**
40					1								1
Lates niloticus	**5**												**5**
41	5												5
Micralestes occidentalis											**3**	**6**	**9**
Oreochromis niloticus												**29**	**29**
Paradistichodus dimidiatus										**11**			**11**
50										11			11
Polypterus endlicheri						**2**							**2**
51						1							1
52						1							1
Polypterus senegalus		**1**											**1**
53		1											1
Raiamas senegalensis									**3**				**3**
54									3				3
Sarotherodon galilaeus												**30**	**30**
Schilbe intermedius			**6**									**5**	**11**
58												5	5
59			6										6
Synodontis nigrita												**31**	**31**
Synodontis puntifer							**1**						**1**
63							1						1
Synodontis schall							**1**					**5**	**6**
62												1	1
63							1						1
64												4	4
Grand Total	**5**	**1**	**6**	**5**	**1**	**2**	**2**	**41**	**3**	**11**	**3**	**215**	**295**

Appendix 2.: Species caught with sequence information, separated by sampling sites and sequencing method.

Table 8: Species table with sampling sites, -areas, catchments and sequence information. Sum of caught fish: number of individuals of certain species. NGS: number of individuals of which a sequence could be retrieved. Sanger of COI: individuals, which were successfully sequenced with Sanger sequencing. SANGER and NGS similar: number of NGS/SANGER sequences that were similar in the COI region. Missing: number of individuals that did not yield either NGS or SANGER data.

Family (N=14)	Genus (N=27)	Species (N=33)	Sampling Catchment (N=2) Sampling Area (N=3) Sampling Site (N=13)													Nr of sequences per species					
			Comoé					**Mouhoun**													
			Banfora					Bobo						Boura							
			Bodadiougou	Diarabakoko	Lake Tangrela	Lembradougou	Moussodougou	Bazudara	Kou	La Genguette	Lake Bala	Naso-Kou	Samandéni-Barrage	Boura-Barrage	Mouhoun River	Sum of caught fish	NGS	NGS of COI	SANGER of COI	SANGER AND NGS similar	MISSING
Alestidae	Brycinus	*B. macrolepidot*											1		2	3	2	0	1		1
		B. nurse							7		10				7	24	24	5	3	0	0
	Micralestes	*M. occidentalis*							7							7	7	6	4	NA	0
		M. sp.													2	2	2	2	1	0	0
	Rhabdalestes	*R. septentrionalis*						12							1	13	13	0	0		0
	Enteromius (former:	*E. macrops*			5	8	1	8	3	1	4		1	10		41	41	3	39	0	0
	Barbus)	*E. sublineatus*							12			2				14	14	1	13		0
	Labeo	*L. senegalensis*					1								5	6	4	3	4	2	2
		L. sp.											1			1	2	1	1	1	0
	Raiamas	*R. senegalensis*										3				3	3	0	3		0
Alestidae sp.													2			2	1	1	0		1
Mormyridae	Hippopotamyrus	*H. pictus*													2	2	2	1	2	0	0
	Marcusenius	*M. senegalensis*					3									3	3	3	1	0	4[1]
	Petrocephalus	*P. soudanensis*													1	1	1	1	0		0

Mochokidae	Synodontis	*S. membranaceus*												1	1	0	0	0		2^{2}
		S. nigrita								12				18	30	34	6	31	0	2^{3}
		S. punctifer								2					2	2	0	1		0
		S. schall												12	12	14	2	7	0	5^{4}
Cichlidae	Chromidotilapia	*C. guntheri*			6										6	5	5	0		1
	Coptodon	*C. zillii*		12	1		5	13					9		40	39	26	2	1	1
	Hemichromis	*H. bimaculatus*		1	6				1	8			11		27	27	24	12	8	0
		H. fasciatus	8		4	2	3			7			10		34	34	13	12	2	0
	Oreochromis	*O. niloticus*		10			6		1	4			9	5	35	34	29	5	3	3^{5}
	Sarotherodon	*S. galilaeus*		4	6	1				7		4	8	5	35	33	30	2	2	2
Claroteidae	Auchenoglanis	*A. occidentalis*												1	1	1	1	1	1	0
	Chrysichthys	*C. auratus*					10	1						2	13	11	11	9	6	2
		C. sp					1								1	1	1	1	0	0
Distichodontidae	Distichodus	*D. rostratus*												1	1	2	1	1	0	5^{6}
	Paradistichodus	*P. dimidiatus*						4					9		13	13	6	9	5	5
Clariidae	Clarias	*C. anguillaris*							1			7		3	11	10	1	4		1
		C. sp.		1						1					2	2	2	1	0	0
Schilbidae	Schilbe	*S. intermedius*								12					12	12	0	11		3^{7}
Anabantidae	Ctenopoma	*C. kingsleyae*							1						1	4	1	1	NA	3^{8}
Bagridae	Bagrus	*B. bajad*												1	1	1	1	1	0	1^{9}
Polypteridae	Polypterus	*P. endlicherii*												2	2	2	0	2		0
Arapaimidae	Heterotis	*H. niloticus*			1										1	2	1	0		1^{10}
Channidae	Parachanna	*P. obscura*								2					2	2	1	1		1^{11}
Latidae	Lates	*L. niloticus*	1				5								6	5	2	5	0	1
		Total N Species	**33**																	
													Total		411	409	191	191	31	47
		Total N Genus	**27**																	
		Total N Families	**14**																	

[1]:4 samples were taken from museum samples
[2]: 1 sample was taken from museum samples
[3]: 6 samples were taken from museum samples, 2 of which are missing
[4]:7 samples were taken from museum, 5 of which are missing
[5]: 1 sample was taken from museum samples
[6]: 6 samples were taken from museum samples, 5 of which are missing
[7]: 5 samples were taken from museum samples, which are missing
[8]: 6 samples were taken from museum samples, 3 of which are missing
[9]: 1 sample was taken from museum samples, which is missing
[10]: 2 samples were taken from museum samples, 1 of which is missing
[11]: 1 sample was taken from museum samples, which is missing

Appendix 3.: Coordinates of sampling sites

Table 9: Sampling site coordinates.

Site	**N**	**W**	
Lake Bala	11.33.548	4.09.304	
Lake Tangrela	10.38.502	4.50.157	
Moussodougou	10.47.255	4.56.112	
Boura-barrage	11.02.458	2.29.564	
Mouhoun	11.01.190	2.49.296	
Bodadiougou	10.46.087	4.55.381	
Lembradougou	10.43.41	4.49.351	
Naso-Kou	11.12.151	4.26.167	
Bazudara	11.14.245	4.26.118	
Kou (Bobo)	11.13.422	4.26.075	
Diarabakoko	10.28.526	4.46.585	
Samandéni	11.23.212	4.34.131	
La Genguette	11.11.159	4.26.260	

Appendix 4:
List of *Synodontis* species for Burkina Faso and neighbouring countries

Table 10: List of *Synodontis* species in Burkina Faso as well as neighbouring countries that share same river catchments according to fishbase.org, 12/2020.
Only freshwater species are shown. x= occurrence, e= endemic

	Burkina Faso	Ghana	Ivory Coast
S. arnoulti	x	x	
S. bastiani	x	x	x
S. batensoda		x	
S. budgetti			x
S. clarias	x	x	
S. comoensis			e
S. eupterus	x	x	
S. filamentosus	x	x	x
S. koensis			e
S. macrophthalmus		e	
S. membranaceus	x	x	
S. nigrita	x	x	
S. obesus		x	
S. ocellifer	x	x	
S. punctifer			e
S. schall	x	x	x
S. sorex	x	x	
S. velifer	x	x	x
S. violaceus		x	
S. voltae	e		
S. waterloti		x	x
Sum	12	16	9

Appendix 5.: PhyML tree for *Oreochromis niloticus*

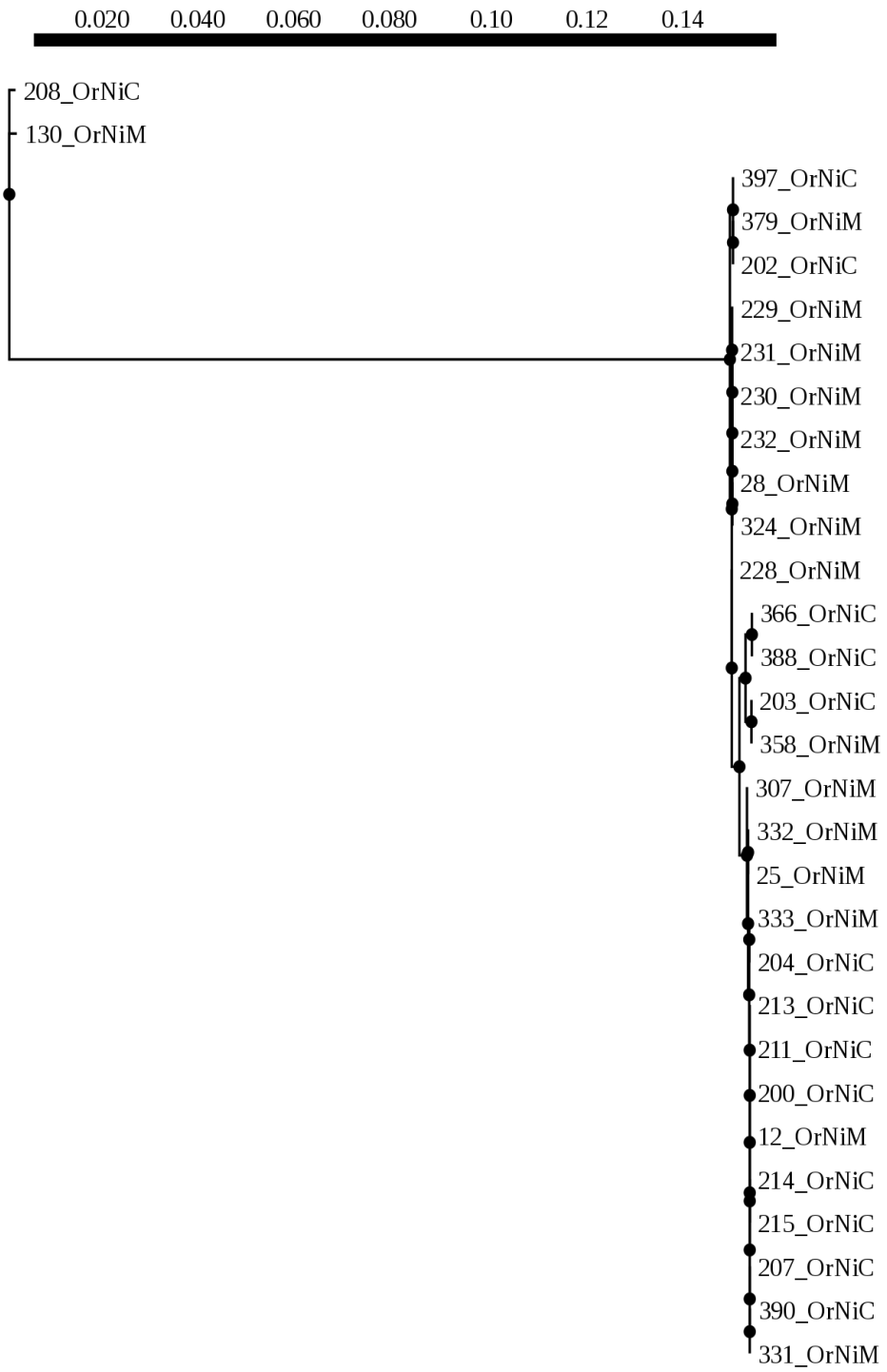

Figure 15.: *Oreochromis niloticus* PhyML Phylogenetic Maximum Likelihood tree. 208_OrNiC: 208: sample number, OrNi: *Oreochromis niloticus,* C: Comoé catchment. 130_OrNiM: 130: sample number, OrNi: *Oreochromis niloticus*, M: Mouhoun catchment

Appenidx 6.: PhyML maximum likelihood tree for the genus *Synodontis*

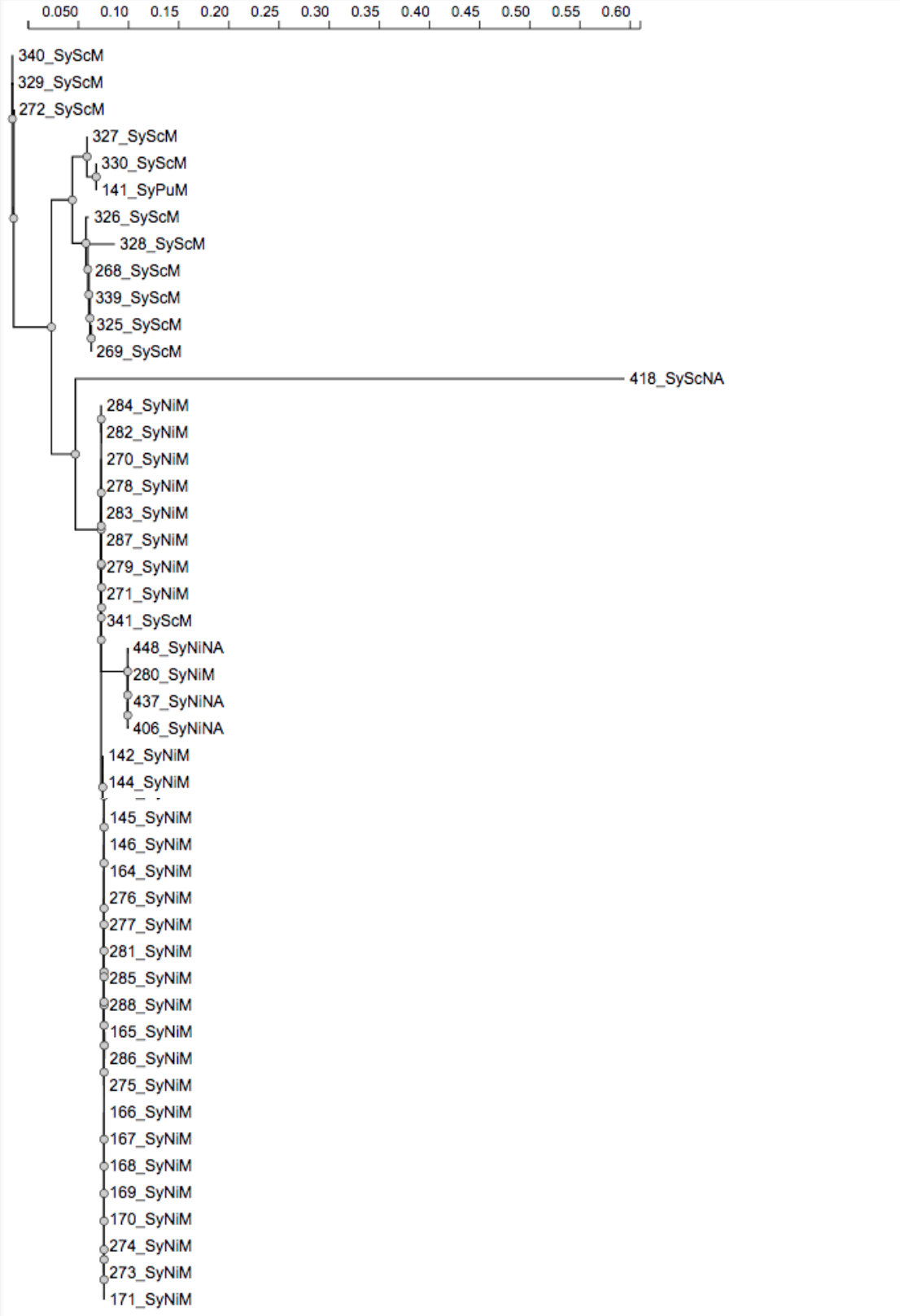

Figure 16.: *Synodontis* PhyML Phylogenetic Maximum Likelihood tree. 340_SyScM: 340: sample number, SySc: *Synodontis schall,* M: Mouhoun catchment. SyNi: *Synodontis nigrita*

Appendix 7.: List of fish species found in Burkina Faso

Table 11.: List of fish species found in Burkina Faso in previous studies (Mano, 2016) compared to species found during this study (2018) and species with barcode records in BOLD

Mano, 2016	Schobesberger, 2018	Record in BOLD
Alestes beramoze		x
Alestes dentex		x
Auchenoglanis occidentalis	x	x
Bagrus bajad	x	x
Bagrus docmak		x
Brienomyrus niger		
Brycinus leuciscus		
Brycinus longipinnis		
Brycinus macrolepidotus	x	x
Brycinus nurse	x	x
Chelaethiops bibie		x
Chrysichthys auratus	x	x
Chrysichthys nigrodigitatus		x
Citharinus citharus		x
Clarias anguillaris	x	
Clarias gariepinus		x
Clypeobarbus hypsolepis		
Coptodon zillii	x	x
Ctenopoma kingsleyae	x	x
Ctenopoma petherici		
Distichodus rostratus	x	x
Enteromius ablabes		
Enteromius baudoni		
Enteromius leonensis		x
Enteromius macrops	x	
Enteromius pobeguini		x
Enteromius punctitaeniatus		
Enteromius sublineatus	x	
Gymnarchus niloticus		x
Hemichromis bimaculatus	x	x
Hemichromis fasciatus	x	x
Hemichromis letourneauxi		
Heterobranchus bidorsalis		
Heterobranchus longifilis		
Heterotis niloticus	x	x
Hippopotamyrus paugyi		x

Hippopotamyrus pictus	x	
Hydrocynus brevis		x
Hydrocynus forskahlii		x
Hydrocynus vittatus		x
Hyperopisus bebe		x
Labeo coubie		x
Labeo senegalensis		x
Lates niloticus	x	x
Malapterurus electricus		x
Marcusenius cyprinoides		x
Marcusenius senegalensis	x	x
Micralestes elongatus		
Micralestes occedentalis	x	x
Micralestes pabrensis		
Mormyrops anguilloides		x
Mormyrus hasselquistii		x
Mormyrus macrophthalmus		x
Mormyrus rume		x
Oreochromis niloticus	x	x
Parachanna obscura	x	x
Paradistichodus dimidiatus	x	x
Parailia pellucida		x
Petrocephalus bovei		
Pollimyrus isidori		x
Polypterus endlicherii	x	x
Polypterus senegalus		x
Protopterus annectens		x
Rhabdalestes septentrionalis	x	x
Sarotherodon galilaeus	x	x
Schilbe intermedius	x	x
Schilbe mystus		x
Siluranodon auritus		x
Synodontis batansoda		
Synodontis clarias		x
Synodontis filamentosus		x
Synodontis membranaceus	x	
Synodontis ocellifer		x
Synodontis punctifer	x	
Synodontis schall	x	x
Synodontis velifer		x
Synodontis vermiculata		