KU LEUVEN

FACULTY OF SCIENCE

Cryptic species diversity in *Enteromius* (Cyprinidae) from West Africa

A morphometric and molecular approach

Douwe BOSMANS

Promotor: Prof. Dr. Jos Snoeks
Ichthyology Department, Royal Museum for Central Africa, Tervuren, Belgium

Co-promotor: Dr. Arthur Boom
Ichthyology Department, Royal Museum for Central Africa, Tervuren, Belgium

Supervisor: Heleen Maetens
Ichthyology Department, Royal Museum for Central Africa, Tervuren, Belgium

Thesis presented in fulfillment of the requirements for the degree of Master of Science in Biology

Academic year 2022-2023

Acknowledgements (Dankwoord)

Deze thesis is tot stand gekomen dankzij de hulp van vele mensen. De steun, begeleiding en aanmoediging die zij tijdens dit proces hebben gegeven is van onschatbare waarde. Daarom neem ik hier even de tijd om ze in de figuurlijke bloemen te zetten.

First and foremost, I would like to thank my promotor Prof. Dr. Jos Snoeks for allowing me to participate in this wonderful research. Your guidance and expertise in African fishes have brought me new insights. A special and heartfelt thank you to my co-promotor Dr. Arthur Boom and supervisor Heleen Maetens. You were always there for me to dance along during the exciting findings, but also to look positively ahead when obstacles appeared. Your guidance and insightful feedback have been invaluable to this research. Thanks to the time and effort you invested in me, I was able to push my limits and develop into a true *Enteromius* researcher. Also a thank you to Dr. Genevieve Diedericks for taking the time to inform us her genetic findings on formalin-fixated specimens. In addition, I would also like to thank the other colleagues of the Ichthyology department for answering my questions and for the nice atmosphere they brought to Tervuren.

Mijn ouders wens ik te bedanken voor hun liefde en steun. Dat vormt het stevigste fundament van mijn succes. Jullie geloofden in mij en waren altijd geïnteresseerd in de vorderingen van de thesis. Ook mijn zus Sien en broer Joos wil ik in dit dankwoord niet overslaan omdat ze zorgden voor de perfecte afleiding en het nodige entertainment om deze thesis af en toe te kunnen vergeten.

Tot slot een luide dankjewel voor mijn vriendin Lene. Hoewel ik af en toe overdonderd werd door dit werk, was jij altijd in de buurt. Jouw aanmoedigingen en geloof in mij hebben me voortgestuwd in de meest uitdagende tijden. Bedankt om aan mijn zijde te staan en mijn trouwe cheerleader te zijn tijdens het hele thesisproces.

Ik ben jullie allen oprecht dankbaar voor de steun en begeleiding tijdens dit intense traject. Jullie hebben uitgedaagd én zijn altijd in mij blijven geloven. Zo werd deze thesis een mooi geheel en leerde ik ook heel wat nieuwe vaardigheden die me zullen helpen in de toekomst.

Douwe Bosmans

Index

Abbreviations

General:

% SL	Percentages of standard length
2X PCR	Two-times PCR
ABGD	Automatic Barcode Gap Discovery
ASAP	Assemble Species by Automatic Partitioning
BLAST	Basic Local Alignment Search Tool
BMNH	British Museum of Natural History (London, UK)
bp	Base pair
C-	Negative control of the PCR
c-B#	Negative control of extraction batch nr #
C+	Positive control of the PCR
COI	Cytochrome c oxidase subunit 1
cyt b	Cytochrome b
DB##	Specimens ID nr##
hDNA	Historical DNA
HP4048	DNA extract of *Hydrocynus forskahlii* (Cuvier, 1819) from the HIPE collection
IUCN	International Union for Conservation of Nature and Natural Resources
LPI	Living Plant Index
MNHN	Museum national d'Histoire naturelle (Paris, France)
mtDNA	Mitochondrial DNA
MWU	Mann-Whitney U
PC	Principal compontent
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
RMCA / MRAC	Royal Museum for Central Africa (Tervuren, Belgium)
RMNH	Rijksmuseum van Natuurlijke Historie (Leiden, Netherland)
s.l.	Sensu lato
s.s.	Sensu stricto
TD	Touchdown PCR
TE	Illumina libraries using targeted enrichment
UPGMA	Unweighted average linkage agglomerative clustering method
UPS water	Ultra-pure sterile water
UV-C	Ultraviolet-C light (100nm and 280nm)
WWF	World Wide Fund for Nature

Morphometry

Measurements:

AFL	Anal fin length
AnFBL	Anal fin base length
BarbA	Length of the anterior barbel
BarbP	Length of the posterior barbel
BD2	Body depth
DoFBL	Dorsal fin base length
DoFL	Dorsal fin length
ED	Eye diameter
HD	Head depth
HL	Head length
HW	Head width
IOW	Interorbital width
MnCPD	Minimum caudal peduncle depth
MxCPD	Maxiumum caudal peduncle depth
PcFL	Pectoral fin length
PlvFL	Pelvic fin length
PoAD2	Post-anal distance 2
PoDD2	Post-dorsal distance
PrAD	Pre-anal distance
PrDD	Pre-dorsal distance
PreOcD	Pre-occipital distance
PrOpD	Pre-operculum distance
PrPecD	Pre-pectoral distance
PrPelD	Pre-pelvic distance
SL	Standard length
SnL	Snout length

Meristics:

AFR	Anal fin rays
code_BarbA	Encoding of the anterior barbel
code_BarbP	Encoding of the posterior barbel
CPS	Scales around caudal peduncle
D-LSc	Scales dorsal fin – lateral line
DFR	Dorsal fin rays
GrLow	Gill rakers on lower limb of first gill arch
GrTot	Total number of gill rakers
GrUp	Gill rakers on upper limb of first gill arch
L-BSc	Scales lateral line - belly
L-PSc	Scales lateral line – pelvic fin
LLS	Lateral line scales
LLS_total	Total lateral line scales
PDSc	Pre-dorsal scales
PecFR	Pectoral fin rays
PelFR	Pelvic fin rays

Samenvatting

Het verlies aan biodiversiteit door klimaatsverandering en menselijke invloed gaat snel. De schatting van het aantal soorten is cruciaal voor de conservatie van de biodiversiteit in zoetwater. *Enteromius* Cope, 1867 (Cypriniformes; Cyprinidae) is een genus van kleine, Afrikaanse zoetwaterbarbelen die gekend zijn voor hun hoge diversiteit met 216 erkende soorten. Soorten van *Enteromius* worden gekarakteriseerd door morfologische gelijkenissen en het voorkomen van verschillende genetische lijnen binnen een morfologische afgelijnde soort. Recent onderzoek in het Congo bassin en Oost-Afrika toonde een patroon van pseudo-cryptische diversiteit aan, wat de aanwezigheid van verborgen diversiteit in *Enteromius* bevestigt. In deze studie onderzochten we de pseudo-cryptische diversiteit van twee soortencomplexen uit West-Afrika: *Enteromius ablabes* / *Enteromius* sp. 'ndoubai' en *Enteromius guildi* / *Enteromius* sp. 'snoeksi'. We voerden morfologische analyses uit op 173 specimens van deze vier soorten op basis van 26 metingen en 13 tellingen en analyseerden de data aan de hand van PCA's en Mann-Whitney U testen. Daarnaast werd er een genetisch protocol ontwikkeld. Van de 173 (formol-gefixeerde ethanol-geconserveerde) specimens, gebruikt voor de morfologie, selecteerden we er 60 voor de genetische analyse. We amplificeerden mini-barcodes met behulp van primers die universeel zijn voor vissen en die twee fragmenten in het COI gen (mtDNA) viseren. Deze fragmenten werden gesequencet door middel van Sanger sequencing. Doordat dit een pionierstudie is in het KMMA en het sequencen van gepreserveerde specimens moeilijk is, was het eerste doel om na te gaan of een DNA-extractie en sequencen met Sanger mogelijk is.

Uit de morfologische resultaten werden twee nieuwe soorten gevalideerd: *E.* sp. 'ndoubai' en *E.* sp. 'snoeksi'. Bovendien waren we in staat de gekende geografische distributie van *E.* sp. 'ndoubai' uit te breiden naar de Little Scarcies rivier in Marela (Guinée). De specimens uit Kalaban (Mali) waren afwijkend ten opzichte van de andere onderzochte specimens, we gaven deze de naam *E.* sp. 'Kalaban'. De erkende soorten *E. ablabes* en *E. guildi* waren in de morfologische analyses moeilijk van elkaar te onderscheiden, wat op een mogelijke synonymie kan wijzen. Bijkomstig genetische analyses zullen echter meer inzicht moeten geven. De DNA-extractie van de formol-gefixeerde ethanol-geconserveerde specimens hadden in het algemeen een lage DNA-concentratie, maar vielen nog steeds in het bereik van bruikbaar DNA volgend de literatuur. De universele primers voor de mini-barcodes in combinatie met de Sanger sequencing bleek echter gevoelig te zijn aan contaminatie van exogeen DNA. Omdat een morfologische aanpak alleen niet voldoende is om de geobserveerde diversiteit te begrijpen, zal toekomstig onderzoek moeten inzetten op een meer geoptimaliseerd genetisch protocol om deze verwachte, verborgen diversiteit in *Enteromius* uit West-Afrika te bevestigen.

Summary

The current rate of species and biodiversity loss is far exceeding the background estimates. The estimation of the number of species is crucial for the conservation measures for freshwater biodiversity. *Enteromius* Cope, 1867 (Cypriniformes; Cyprinidae) is a genus of African freshwater minnows known for its enormous amount of diversity, with already 216 valid species reported. Species of *Enteromius* are characterized by morphological similarities. Additionally, genetic evidence suggests the occurrence of highly different genetic entities in several morphological species. Case studies in the Congo basin and East Africa already showed a pattern of pseudo-cryptic diversity, suggesting that the hidden diversity is prominent in the *Enteromius* genus. In this work, we explored the pseudo-cryptic diversity in West Africa, by examining two species complexes: *Enteromius ablabes* / *Enteromius* sp. 'ndoubai' and *Enteromius guildi* / *Enteromius* sp. 'snoeksi'. First, we examined 173 specimens of the four species in a morphometric approach using 26 measurements and 13 meristics. We interpret the data performing PCAs and Mann-Whitney U tests. Second, we conducted genetic analysis on a subsample of those specimens (formalin-fixed ethanol-preserved). Of the 173 specimens used in the morphometric approach, 60 were selected for the genetic approach. We attempted to amplify mini-barcodes using primers universal for fish targeting two fragments of the COI gene (mtDNA) and Sanger sequencing. As it is a pioneer study in the RMCA and as genetic methods on such samples are known to be challenging, our primary objective was to determine the feasibility of DNA extraction and Sanger sequencing on these specimens.

We validated two new species, *E.* sp. 'ndoubai' and *E.* sp. 'snoeksi' using the morphometric approach. Furthermore, we updated the known geographic distribution of *E.* sp. 'ndoubai', known from the Kakrima River (Konkouré basin in Guinea) with records in the Little Scarcies River in Marela (Guinea). We also found outliner specimens that we called *E.* sp. 'Kalaban' from Kalaban (Mali). The valid species *E. ablabes* and *E. guildi* were difficult to distinguish based on morphology, which raised the question of a putative synonymy. However, genetic analyses should give more insights. The DNA extractions conducted on the formalin-fixed specimens have overall low DNA concentrations but still fall within the range of usable DNA according to the literature. However, the universal mini-barcode primers tested in this work in combination with Sanger sequencing appeared to be really sensitive to exogenous DNA contamination and we were not able to get mini barcode sequences. As the morphometric approach alone is insufficient to understand properly the observed diversity, future work using other sequencing methods should greatly help to assess the expected hidden diversity in *Enteromius* from West Africa.

1. Introduction

1.1. Biodiversity

The word “biodiversity” was originally used by Dr. Walter G. Rosen in 1986 at the National Forum on BioDiversity in Washington, D.C. to describe the enormous diversity of the living world (National Academy of Sciences, 1988). In 1992, the Convention on Biological Diversity of the United Nations defined “biodiversity” as “the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are a part; this includes diversity within species, between species and of ecosystems”. Biological diversity receives great attention because it is considered a global resource to be preserved (Lévêque, 1994; Ceballos et al., 2015). Understanding biodiversity and its characterisation is essential because of its role in regulating ecosystems on which humans depend (Lévêque, 1994; Teugels & Hopkins, 2007; Bamba, 2012).

1.1.1. The current loss in biological diversity

All around the world, a drastically accelerated transformation of natural landscapes is taking place with the associated loss of species due to anthropological activities (Lévêque, 1994; WWF, 2022). The current rate of species losses is far exceeding the background estimates, resulting in the hypothesis of a sixth global mass extinction event (Barnosky et al., 2011; Kolbert, 2014; Ceballos et al., 2015; Cowie et al., 2022). Compared to the other five mass extinction events in Earth’s history, the Sixth Mass Extinction is caused entirely by humans (Cowie et al., 2022). According to Ceballos et al., (2015), the average rate of vertebrate species loss over the past century is 100 times greater than the pre-anthropogenic background rate, while using cautious assumptions. Even the rate of invertebrate extinctions is estimated to vastly exceed the background rate (Cowie et al., 2022). Even though there is still debate around this idea of a sixth global mass extinction event, the current biodiversity loss cannot be ignored. The loss of biodiversity due to human influences concerns terrestrial and marine ecosystems, but also freshwater systems in a multileveled and complex way. Just 0.01% of the world’s water and approximately 2.3% of Earth’s surface is fresh water; yet this tiny fraction hosts almost 9.5% of all described animal species (Dudgeon et al., 2006; Reid et al., 2019). According to the 2022 Living Plant Index (LPI), populations of freshwater species have decreased on average by 83% since 1970, which is a much greater loss than for terrestrial or marine species (Harrison et al., 2018; WWF, 2022).

1.1.2. Threats to freshwater biodiversity

Anthropogenic influences like overexploitation of resources, pollution, deforestation and species introductions are causing a rapid change in climates, resulting in a loss in biodiversity on a global scale (WWF, 2022). For freshwater ecosystems, human activities could be classified into five major categories (Lévêque, 1994; Dudgeon et al., 2006; Thieme et al., 2006; Bamba, 2012; Darwall et al., 2018). First is the pollution of aquatic systems by agricultural or industrial activities, like extensive nutrient enrichment and microplastics (Dudgeon et al., 2006; Reid et al., 2019). The second category is the competition bound in the search for water (Lévêque, 1994, 1997). Furthermore, the degradation and destruction of habitats through streams and river canalization and dam construction have a major impact on freshwater diversity (Dudgeon et al., 2006; Reid et al., 2019; King & Brown, 2021). Hydropower and irrigation dams in the river systems are stopping the upstream and downstream movement of species, meaning spawning sites cannot be reached, resulting in a decline in species and habitat diversity (King & Brown, 2021). The introduction of exotic species into freshwater systems, e.g. aquaculture purposes, is listed as the fourth category (Lévêque, 1997; Dudgeon et al., 2006). As the last major category, there is the overexploitation and extirpation of species due to intensive fisheries for consumption (Dudgeon et al., 2006; Reid et al., 2019; WWF, 2021). Compared to the other four categories, which affect also microfauna, unsustainable fishing affects mainly macrofauna like fish, amphibians and reptiles (Dudgeon et al., 2006). Besides these five anthropogenic categories, there are also new emerging threats to freshwater biodiversity like climate change, infectious diseases, harmful algal blooms, nanomaterials, light and noise, cumulative stressors and more (Reid et al., 2019). Climate change, resulting in amplified climate extremes, has a larger effect on the freshwater habitats in the tropics than on those in temperate zones (Barbarossa et al., 2021).

According to the “World’s Forgotten Fishes” report produced by WWF and 15 other organizations, freshwater fish species are particularly endangered, with up to one-third of them being threatened with extinction (WWF, 2021). Conservation of freshwater ecosystems often comes second to other development priorities that are seen as more directly linked with human well-being, e.g. drinking water, energy, and food (Darwall et al., 2018). Additionally, to have an effective way of freshwater conservation, requires control over the upstream drainage network, the surrounding land, the riparian zone and downstream reaches. Such requirements are rarely realized; hence the creation of inclusive and international management partnerships at suitable (drainage-basin) scales will be necessary (Dudgeon et al., 2006). There have been policy initiatives that aim to protect freshwater life and ecosystems. Still, these are often not sufficient and not carried out correctly (Dudgeon et al., 2006; Darwall et al., 2018).

1.1.3. The lack of taxonomic knowledge

The loss of biological diversity and the importance of the conservation of freshwater habitats are likely to be underestimated as a result of the Linnean Shortfall and the cryptic species diversity (Bickford et al., 2007; Adams et al., 2014; Poulin & Pérez-Ponce de León, 2017; Struck et al., 2018; Walters et al., 2021). In attempts to estimate the true number of species on Earth, there are several shortfalls. Hortal et al., (2015) explain seven shortfalls of biodiversity knowledge for evolutionary and ecological research. The Linnean Shortfall refers to the difference between the number of formally described species and the actual number of species on Earth (Hortal et al., 2015). This shortfall is one of the most notable as it affects the estimations of the actual number of species on Earth (Hortal et al., 2015). Furthermore, the Linnean Shortfall interferes and is connected with the other shortfalls explained by Hortal and colleagues (2015) and is a widely known pitfall in the estimations of the number of species. The magnitude of this shortfall is still unknown because (i) there are continually new species being formally described and (ii) difficulties in establishing a unified species concept (Hortal et al., 2015; Walters et al., 2021).

Furthermore, cryptic species diversity affects this Linnean Shortfall and thereby also the estimation of the total number of species on Earth (Walters et al., 2021). Due to the greater habitat heterogeneity and geographical isolation, freshwater ecosystems exhibit higher frequencies of cryptic taxa than marine or terrestrial ones (Poulin & Pérez-Ponce de León, 2017). The concept of ‘cryptic species’ is frequently used in studies in a variety of ways to label species that are difficult to recognise as distinct species, and some scientists question the biological relevance of it (Heethoff, 2018; Struck et al., 2018; Pfingstl et al., 2021). Moreover, numerous variations are used across the literature, such as “hyper-cryptic” (e.g. by Adams et al., 2014) and “pseudo-cryptic” (e.g. by Luttikhuizen & Dekker, 2010) (Struck et al., 2018). These inconsistent definitions and taxonomic treatments of cryptic species cause much debate among researchers and make the comparisons between the studies more complex (Struck et al., 2018; Pfingstl et al., 2021). Adams and colleagues (2014) defined ‘hyper-cryptic species complexes’ as: “any taxon currently regarded as a single species or any related group of taxonomically confused species, that in reality consists of a “large” number of valid but undiagnosed species.” In this dissertation, we use the concept ‘cryptic species’ as two or more species have been classified as a single nominal species because they are morphologically indistinguishable, yet evidence indicates they are on different evolutionary trajectories, following Bickford et al., (2007) and Struck et al., (2018) and ‘pseudo-cryptic species’ as species that show minor morphological differences that are only detected by detailed examination, following Luttikhuizen & Dekker (2010).

1.2. Biological diversity of fishes in Africa

The diversity of freshwater life is abundant throughout the African continent, from wet, tropical regions to the more temperate parts. Compared to communities in temperate zones, tropical communities of plants and animals are characterised by their high species diversity and intricate interrelationships (Lowe-McConnell, 1987). Tropical regions are known for their seasonal variation by wind regimes and fluctuations of rainfall. Fluctuating levels of rainfall throughout the year lead to regular flooding and expand the freshwater environment seasonally, causing variations in most tropical waters (Lowe-McConnell, 1987).

The rivers and lake basins of Africa are among the oldest in the world and have a much longer history than those in the temperate regions (Thieme et al., 2006). More phylogenetically isolated freshwater fishes are found in Africa than on any other continent, and there have been exceptional species radiations among many different taxa in rivers and lakes (Roberts, 1975; Lowe-McConnell, 1987; Thieme et al., 2006). The African continent can be divided into two major parts based on freshwater fishes: High Africa (elevations mainly above 1000 m) and Low Africa (150-600 m elevations) (Roberts, 1975). West Africa lies completely in Low Africa and almost all of the phylogenetically isolated freshwater fishes can be found in Low Africa (Thieme et al., 2006). Species are not evenly distributed in Africa because of tectonic and geological events and the subsequent diversifications. By using the bioregionalization method, scientists could thereby distinguish geographic regions that have their own set of faunal and floral species and differ by species composition (Thieme et al., 2006; Leroy et al., 2019). In 1975, Roberts defined ten ichthyofaunal provinces on the African continent based on the phylogenetic history, palaeogeography and ecology and thereby laying the foundations of the ecoregions defined today (Thieme et al., 2006; Abell et al., 2008).

We define ecoregions as “a large area of land or water containing a distinct assemblage of natural communities and species, whose boundaries approximate the original extent of natural communities before major land use changes” and bioregions or biogeographic provinces as “a complex of ecoregions that share a similar biogeographic history and thus often have strong affinities at higher taxonomic level” as in Thieme et al., (2006). To date, Africa and Madagascar can be divided into a set of eleven bioregions and 93 ecoregions (Thieme et al., 2006). Of the 60.000 vertebrate species described, more than one-half are teleost fish, with 3680 species inhibiting the inland waters of Africa. (Nelson et al., 2016; Froese & Pauly, 2022). The African species diversity is probably underestimated and might either correspond to regional fauna yet to be properly understood (e.g. Great African lakes, Congolian and Angolan river systems in Snoeks et al., (2011)) or to some taxa known to host potential cryptic species (cf. 1.2 section) (Decru et al., 2016; Van Ginneken et al., 2017).

The present-day distribution of fish is the result of past climate (e.g. alternating wet and dry periods) and geological (e.g. the appearance of dispersal barriers) events as these barriers were uncrossable for almost all living fish species (Daget & Iltis, 1965; Hugueny & Lévêque, 1994).

1.2.1. Ecoregions and ichthyo-diversity in West Africa, the study area

In 1965, Daget & Iltis determined three geographical categories in West African freshwater fishes: Sudanese, Guinean and indifferent. Later, a revision of these findings was done by Hugueny & Lévêque (1994), resulting in three main zoogeographic regions: The Sudanian region sensu stricto, (s.s.) the Upper Guinean region, and the Eburneo-Ghanean region, which overlapped with most of the regions in Daget & Iltis (1965). Twenty-one to 71% of the fish species in each of these zoogeographic regions are endemic, although some species entered the different basins by migrating through brackish waters (Hugueny & Lévêque, 1994).

More recently, for West Africa, Thieme et al., (2006) defined almost twenty ecoregions falling into three major bioregions (Nilo-Sudan, Upper Guinea and West Coast equatorial) (Figure 1). This part of the continent has a variety of habitat types including moist forest rivers, savanna-dry forest rivers, large river deltas and some highland and mountain systems presented (Thieme et al., 2006). The moist forest rivers are characterised as the most species-rich compared to the other habitat types in Africa. Thereby Western Africa, together with the Congo basin and the Rift Vally Lakes, has the highest species numbers on the continent (Thieme et al., 2006). The larger study basins in West Africa that we include in this thesis are the Konkouré Basin in Guinea, the Comoé, Tanoé and Cavally basins in Ivory Coast and Ghana and the Middle Niger basin in Nigeria (Figure 1). A more complete overview will be given in the Material and Methods part of the thesis.

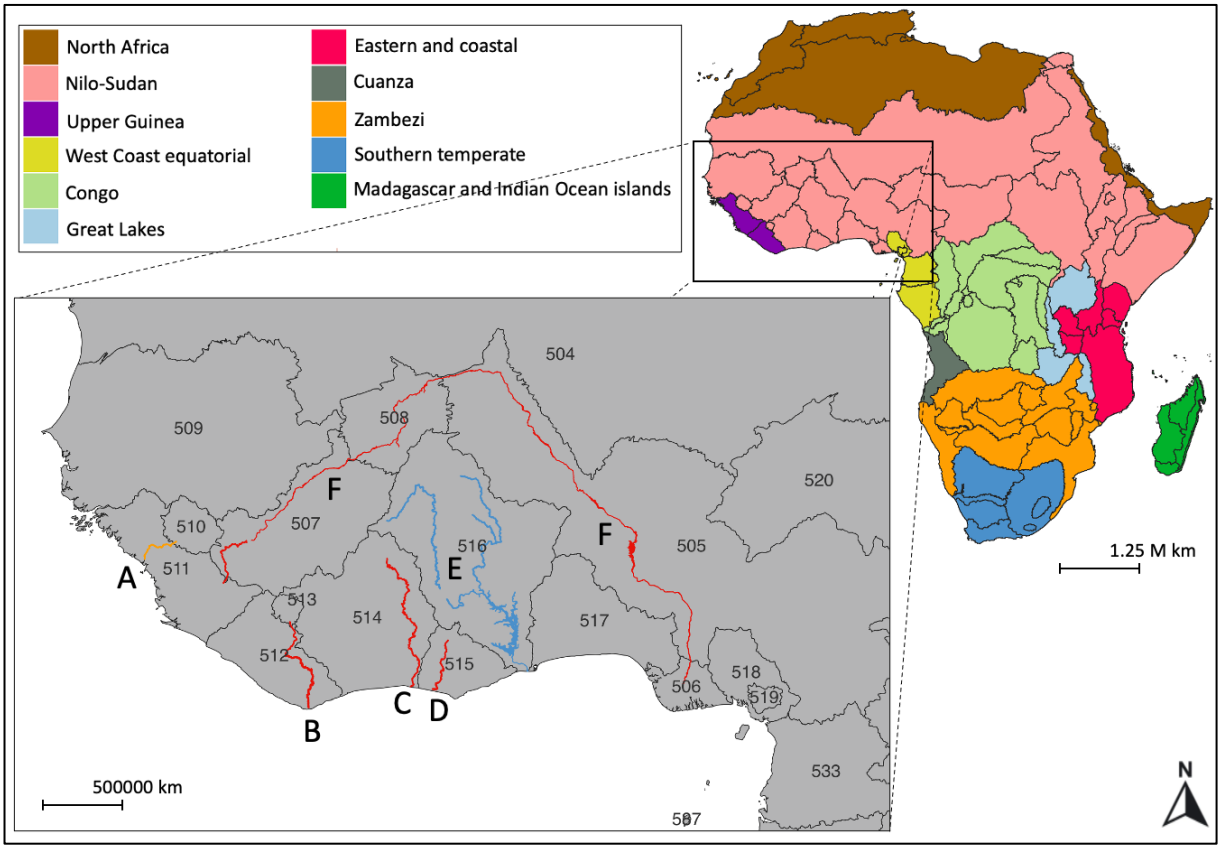

Figure 1: Map of the African bioregions explained by Thieme et al., (2006) including the ecoregions and most notable river systems, for this thesis, of West Africa. The numbers presented in West Africa are the ecoregions, which can be found in Addendum B. A) is the Konkouré River, B) the Cavally River, C) the Comoé River, D) The Tanoé River, E) the Lake Volta system including rivers, F) the Niger River. This map is made in QGIS using the shape files of Abell et al., (2008) for bioregions and ecoregions and the shape files from DIVA-GISS for the river systems.

Even if there is a rise in attention to the loss of biodiversity in tropical rain forests or coastal areas, the diversity of and within freshwater systems has been mostly ignored (Lévêque, 1994). The unsatisfactory taxonomic knowledge of the African ichthyofauna makes the conservation of freshwater ecosystems even harder (Decru et al., 2016; Van Ginneken et al., 2017). This holds for West African freshwater ecosystems, that face a lack of taxonomic knowledge regarding the fish diversity (Darwall et al., 2018; Schmidt et al., 2019). The characterization of species that remains to be discovered and of cryptic diversity are conclusive of critical importance. Indeed, incomplete inventories can lead to incorrect management and conservation policy, and an underestimation of the magnitude of human impact on the rate of species loss (Decru, 2015; Decru et al., 2016; Darwall et al., 2018). Indeed, This could result in the loss of biodiversity and productivity, but also in the loss of essential ecosystem services.

1.1.1. The family Cyprinidae

Cyprinidae is a family within the order Cypriniformes. This family includes 1780 valid species, in 157 valid genera, making it the largest family of freshwater fish (Fricke et al., 2022; Froese

& Pauly, 2022). Cyprinids are characterised by a protractible mouth with toothless oral jaws, a toothless palate and well-developed pharyngeal teeth (Howes, 1991; Nelson et al., 2016). The pelvic fins have an abdominal position and no adipose fin is present (Fricke et al., 2022). Some species within Cyprinidae have barbels and/or spinelike rays in the dorsal fin (Nelson et al., 2016). Sexual dimorphism can occur in colour patterns and through breeding tubercles or pearl organs in males (Manda et al., 2020). Representatives of this family have a widespread distribution and can be found in almost all kinds of freshwater habitats throughout Eurasia and Africa and certain parts of North America (Howes, 1991; Nelson et al., 2016). Although some species can survive and even breed in brackish water, most species are restricted to freshwater habitats (Howes, 1991). The Cyprinidae are often used in aquaculture (e.g. Common Carp, *Cyprinus carpio*), as aquarium fish (e.g. Goldfish, *Carassius auratus*), for biological, ontogenetic, and genetic research (e.g. zebrafish, *Danio rerio*) (Nelson et al., 2016). As the Cyprinidae family is frequently associated with numerous identification issues, the phylogenetic relations within the group are being studied more and more (Bamba, 2012).

1.2. *Enteromius* Cope (1867)

More than 800 nominal species, distributed in Eurasia and Africa, were assigned to the genus *Barbus* sensu lato which was introduced by Cuvier and Cloquet in 1816 based on several morphological characteristics like the size and placement of the mouth, the size and placement of the fins and the number and arrangement of scales (Berrebi et al., 1996). Cuvier and Cloquet (1816) distinguished several species within *Barbus* based on differences in external features, including colouration. Numerous studies suggested that such a wide distribution and large range of morphologic differences within one genus indicated a polyphyletic assemblage of several unrelated lineages (Myers, 1960; Howes, 1987; Agnèse et al., 1990; Skelton et al., 1991; Berrebi et al., 1996; Yang et al., 2015). Early revisions on *Barbus* sensu lato revealed three groups according to their ploidy level: diploids (2n = 48 or 50), tetraploids (2n = 100) and hexaploids (2n = 148-150) (Agnèse et al., 1990; Berrebi et al., 1996; Mullens et al., 2020). Based on this ploidy level, *Barbus* s.l. was split into several genera, e.g. the genus of the “true barbs”, *Barbus* sensu stricto (s.s), which includes the Mediterranean tetraploids barbs (Berrebi et al., 1996; Berrebi et al., 2014). Species of *Barbus* s.l. that could not be assigned to a proper genus were assigned to ‘*Barbus*’ (between single quotation marks) (Berrebi et al., 1996). Large hexaploid African barbs were placed into *Labeobarbus* Rüppell (1835) (Agnèse et al., 1990; Berrebi & Valiushok, 1998). Yang and colleagues (2015), recently did a taxonomic revision, including the genus *Barbus* s.l., using genetic evidence (both nuclear and mitochondrial DNA sequences). They established the relationships of cyprinid fish (i.e. phylogeny) and proposed to accommodate all small-sized African diploid ‘*Barbus*’ in the genus *Enteromius* Cope, 1867. A recent study by Hayes & Armbruster (2017) insisted on three certainties about *Enteromius*:

They are not *Barbus*, not even closely related to *Barbus*, and will never be placed back in *Barbus*. *Enteromius* in a non-monophyletic group as some genera like *Caecobarbus*, *Barboides* and *Prolabeops* are nested within the *Enteromius* clades, based on mtDNA (Yang et al., 2015; Hayes & Armbruster, 2017; Mullens et al., 2020; Schedel et al., 2022).

Enteromius includes 226 valid species (Froese & Pauly, 2022). Species within this genus have generally seven to eight branched dorsal fin rays, moderately developed lips and poorly developed gill rakers (Lévêque, 1983; Lévêque et al., 1990; Bamba, 2012). One or two pairs of barbels are present (Lévêque, 1983). The colouration of the representatives of the genus is plastic and varies between and within populations (Lévêque et al., 1990). They have a pan-African distribution and are restricted to the continent (Skelton et al., 1991)

1.2.2. Taxonomic uncertainties in *Enteromius*

Apart from the taxonomic difficulties discussed in the previous sections, several genetics-based studies (Decru, 2015; Schmidt et al., 2017; Van Ginneken et al., 2017; Maetens et al., 2020; Decru et al., 2022) suggest that the species diversity is probably dramatically underestimated, with many pseudo-cryptic and cryptic species. Studies on focal regions already exposed strongly different genetic groups, delineated using phylogenetic approaches, only slightly different in terms of morphology (Schmidt et al., 2017; Van Ginneken et al., 2017; Decru et al., 2022; Maetens et al., in prep.). The incomplete morphological descriptions of several species make it challenging to identify species in the field (Berrebi et al., 1996; Decru et al., 2016; Maetens, 2019). In addition, literature on *Enteromius* is often limited to the original description as identification keys and reviews for larger regions are lacking (Hayes & Armbruster, 2017; Van Ginneken et al., 2017; Maetens, 2019). This leads to misidentifications in the field and consequently on GenBank (Hayes & Armbruster, 2017). The lack of reference genes, the polytomy, and the paraphyly further exacerbate the problematic situation of the genus (Hayes & Armbruster, 2017; Mullens et al., 2020; Schedel et al., 2022). There is a strong consensus on the urgent need for additional work to clarify the taxonomy of *Enteromius* (Bamba, 2012; Decru, 2015; Hayes & Armbruster, 2017; Schmidt et al., 2017; Van Ginneken et al., 2017; Maetens, 2019; Maetens et al., 2020; Mullens et al., 2020; Decru et al., 2022; Schedel et al., 2022).

An integrative approach using morphology and genetics has been proven to be useful to resolve the complex diversity in small barbs (e.g. Van Ginneken et al., 2017; Maetens et al., 2020). Traditional morphology includes measuring certain structures and counting several meristics. Different elaborate methods of measuring like landmark-based geometric morphometrics and outline-based morphometrics are now often used by taxonomists (Dujardin, 2017). In the previous two decades, the advancement of molecular research has

made it possible to distinguish between species using a wide range of approaches. DNA barcoding on mitochondrial DNA, like cytochrome c oxidase I (COI) and cytochrome b (cyt b), as well as whole genome comparisons, are used as reliable methods to delineate species (Van Ginneken et al., 2017; Piñeros et al., 2022). Molecular and morphology-based techniques are in many cases not sufficient on their own. A combination of both approaches is generally needed to revise problematic taxa and to delineate and describe species (Wheeler, 2004; Decru, 2015).

1.2.3. Case studies on *Enteromius*

As mentioned above, numerous studies have been done to detangle the taxonomic diversity in the genus *Enteromius*.

In the northeastern part of the Congo Basin, Van Ginneken et al., (2017) found 23 mitochondrial (COI) lineages in what was thought to be only four species. After a closer examination with multivariate analyses, also minor morphological differences were found, which suggests the finding of separate, undescribed species (Van Ginneken et al., 2017). Similar results were found in the Lake Edward system, where several species represent multiple genetic lineages (mitochondrial, COI) (Decru et al., 2022; Maetens et al., in prep.). Still, in East Africa, Schmidt et al., (2017) also characterize unrecognised diversity within the species of *Enteromius* present all across Kenya by using multiple genetic markers.

Most of the aforementioned cases occur in East and Central Africa. Still, we can expect a similar pattern for *Enteromius* in other regions. In Western Africa, Schmidt et al., (2019) studied *Enteromius* in the Fouta Djallon highlands in the Republic of Guinea. When studying *Enteromius foutensis* (Lévêque, Teugels & van den Audenaerde, 1988), three divergent and well-supported populations of the species were observed based on a molecular analysis with cyt b and RAC1 sequences. From these lineages, they recognised two candidate species distinct from *E. foutensis* when combining both molecular and morphological approaches (Schmidt et al., 2019). More recent work also suggests unrecognized diversity of *Enteromius* in the Rover River basin in Sierra Leone (Kanu et al., 2022). Kanu and colleagues (2022) found, based on cyt b (mtDNA) sequences, numerous possible new species in *E. foutensis, Enteromius ablabes* (Bleeker, 1863) and *Enteromius guildi* (Loiselle, 1973). As the study is conducted on a limited scale (i.e. Rover River basin), we might expect more unrecognised diversity in the aforementioned species at the whole Western Africa scale. Interestingly, a revision (Bamba, 2012) conducted on *'Barbus'* from the Ivory Coast (and surroundings) already recognized potential new species (*Enteromius ndoubai* sp. nov. *and Enteromius snoeksi* sp. nov.) for the last two mentioned species, *E. ablabes* and *E. guildi*.

Enteromius ablabes (Figure 2A) is a small barb known from the Sahelian as well as from coastal basins and some populations in the Central African Republic (Froese & Pauly, 2022). Lévêque (1990) stated that *E. ablabes* could be a complex of species or a polymorphic species. This statement on polymorphism was rejected by Bamba (2012) as he never found two discrete morphotypes occurring together in the same population. He examined specimens from three different counties: Côte d'Ivoire, Ghana and Guinea. Based on morphometric analyses, he found that the specimens from the Kakrima River in the Konkouré Basin of Guinea were separated from other populations. The Kakrima River is known for its high number of endemic species due to the isolation of large parts by e.g. waterfalls (Bamba, 2012; Schmidt, 2014). Bamba (2012) observed longer barbels, absent or hardly visible cephalic canals, deeper and shorter caudal peduncle, a lower number of lateral line scales and a lower number of total gill rakers in specimens of the Kakrima river, compared to other populations of *E. ablabes*. He assigned specimens from the Konkouré basin to a new species: *Enteromius ndoubai* sp. nov. (Figure 2B). Because the species has not been described, we refer to this species below as *Enteromius* sp. 'ndoubai'.

As mentioned, Bamba (2012) also studied *E. guildi* from Ivory Coast (Figure 2C). *Enteromius guildi* is morphologically very similar to *E. ablabes* but can be differentiated by 10 scales around the caudal peduncle (vs. 12 in *E. ablabes*) and a lateral line that is depressed below the dorsal fin (Lévêque, 2003). However, by comparing fresh specimens from the two species, Bamba (2012) concluded that these differences are contested. He noticed a large overlap for some characters (e.g. in body shape) considered as diagnostic between *E. guildi* and *E. ablabes*, which could point to a possible synonymy of both species. Still, based on previous observations, the two species were considered valid. Bamba (2012) also looked into intra-specific morphological variation within *E. guildi*. Representatives of the species in the Tanoé River (Côte d'Ivoire) could be distinguished from the other populations based on multivariate analyses. The population of the Tanoé River was proposed as a new species, *Enteromius snoeksi* sp. nov. (Figure 2D). Because the species has not been described, we refer to this species below as *Enteromius* sp. 'snoeksi'. This species can be distinguished from other populations of *E. guildi* by a larger eye diameter, a smaller body depth, a larger post-anal and post-dorsal distance, and a smaller snout length (Bamba, 2012).

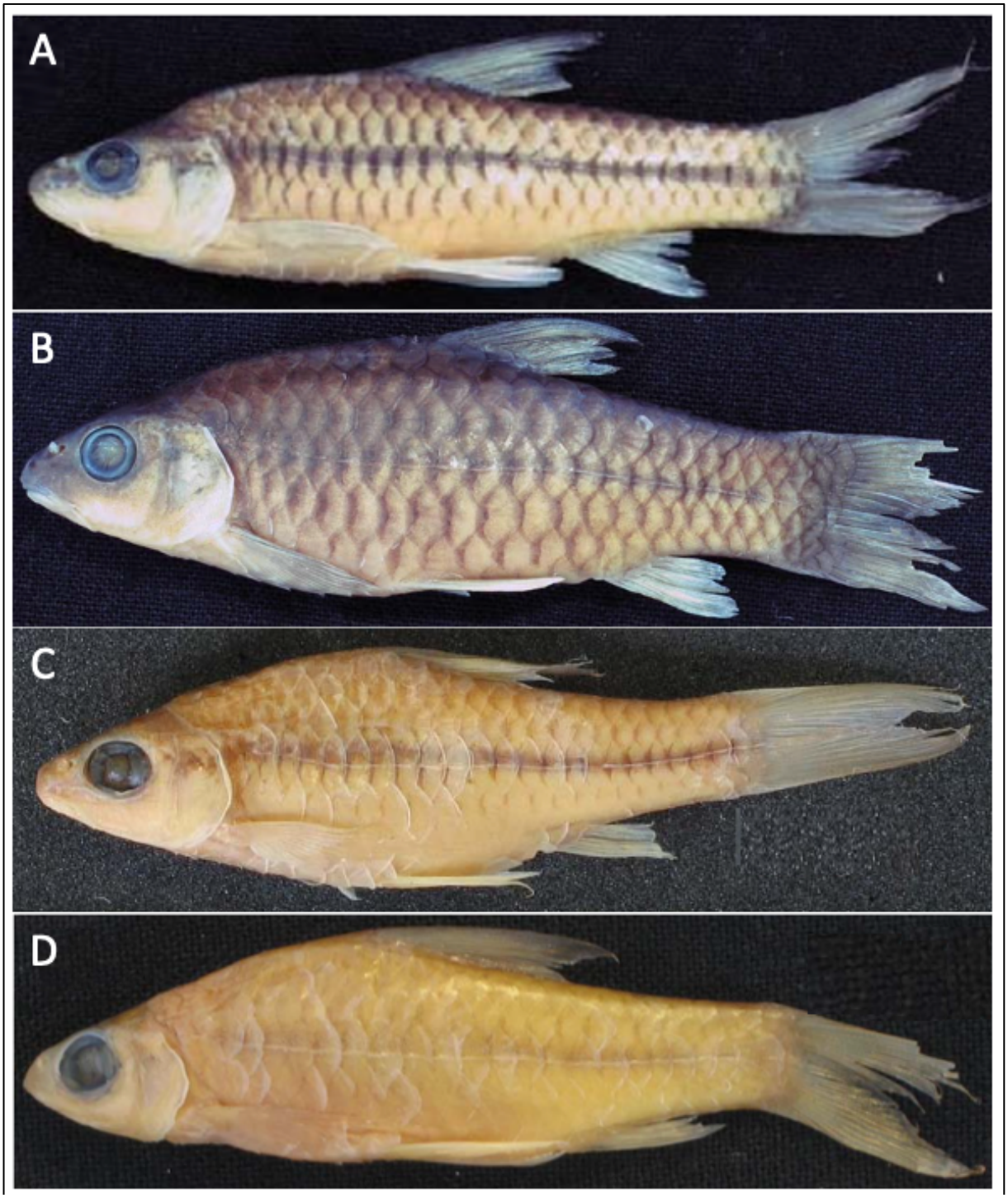

Figure 2: A) RMNH 2466, a syntype of *E. ablabes*, Dabo Krom (Southern Ghana). B) MRAC 92-059-P-0874, holotype of *E.* sp. 'ndoubai', Galekoulou River in basin Konkouré (Guinea). C) BMNH 1971.9.13.3-7, a paratype of *E. guildi*, Hedjo River in Volta basin (Ghana). D) MRAC A5-43-P-165, holotype of *E.* sp. 'snoeksi', Tanoé River in basin Tanoé (Côte d'Ivoire). Pictures form Bamba (2012).

1.3. Species delineation on museum specimens

1.3.1. DNA extraction of historical DNA

To preserve specimens for long-term purposes, the most common method involves several steps, including fixation of the specimens with formalin on the field (Hykin et al., 2015; Appleyard et al., 2021). Historical DNA (hDNA) refers to the extracted DNA from preserved biological samples, such as museum specimens. The use of formalin has many negative impacts on the DNA as these formalin fixatives lead to DNA degradation and fragmentation, as well as DNA-protein cross-linking and base modification (Hykin et al., 2015; Appleyard et al., 2021). These impacts limit the molecular usage of preserved specimens due to the extremely low availability of usable, non-degraded DNA (Schander & Halanych, 2003; Garrigos et al., 2013; Hykin et al., 2015; Appleyard et al., 2021; Hahn et al., 2022). Still, using a museum collection holds enormous potential as it allows to include of a wide and unparalleled amount of data that are otherwise difficult to acquire (Raxworthy & Smith, 2021). Different protocols and techniques are developed, but still, extracting DNA is difficult (Chakraborty et al., 2006). More recently, some researchers were able to use museum specimens for molecular studies (Garrigos et al., 2013; Hykin et al., 2015; Diedericks, 2017; Appleyard et al., 2021; Hahn et al., 2022), but this was mostly done in special ancient DNA laboratories. In this thesis, we will test the feasibility of DNA extraction and sequencing from formalin-fixed, ethanol-preserved museum specimens in the RMCA in Tervuren.

1.3.2. DNA barcoding

Hebert et al., (2003a) showed the possibility of using the mitochondrial gene cytochrome c oxidase subunit 1 (COI) to explore diversity and delineate species. They showed that closely related species in all animal phyla, except the Cnidaria, can be discriminated by using COI sequence divergences as this gene evolves rapidly (Hebert et al., 2003a, 2003b). The COI gene can be regarded as a “barcode” for a species, meaning that the divergence observed in this gene is a result of speciation. Other barcodes can also be used such as cytochrome b (cyt b), ribosomal S16 gene and ribosomal S18 gene (Tang et al., 2012; Schmidt et al., 2019).

DNA barcoding was used in the first place to identify specimens to the species level but is now also used for the delineation of species (Hebert et al., 2003b; Van Ginneken et al., 2017). In many cases (e.g. Adams et al., (2014) and Zamani et al., (2021)), DNA barcoding as a stand-alone method provides insufficient evidence for accurate species delineation. DNA barcoding should then be combined with other approaches, such as the use of microsatellite genotyping as in Hubert & Hanner, (2015) or morphological analyses as in Van Ginneken et al., (2017), for proper species delineation. To investigate hidden species diversity, the sequencing of mitochondrial DNA barcodes, like COI and cyt b, is supported by complementary properties

such as the barcoding gap (Puillandre et al., 2012). It is expected that the distribution of pairwise differences across sequences will rely on both intraspecific and interspecific diversity when dealing with a barcode dataset that includes several species. Intraspecific distances are theoretically lower than interspecific distances. Both are typically separated by a hiatus or the barcode gap. Such gaps are used by several techniques (e.g. Automatic Barcode Gap Discovery (ABGD), Assemble Species by Automatic Partitioning (ASAP), etc.) to indicate the division between putative different species (Oh et al., 2022). These methods can be complemented with maximum-likelihood or Bayesian methods in software like RaxML[1] and MrBayes[2] to estimate the relationship between sequences and thereby provide a proxy of the species genealogy (Maddison, 1997). Based on such genealogy, it is possible to identify clusters corresponding to different species. Still, due to deep coalescence, lack of polymorphisms and hybridization, a phylogenetic diagram based on a limited number of genes might poorly represent the species tree (Maddison, 1997; Hubert & Hanner, 2015).

1.3.3. DNA mini-barcoding

As we are working with formalin-fixed specimens, it may be difficult to retrieve enough DNA of sufficient quality to be able to amplify classic barcodes. A suitable alternative consists of performing shallow whole-genome sequencing, with appropriate kits for both DNA extractions and library preparations, and concentrating on the assembly of organelle genomes or repetitive DNA sequences (Straub et al., 2012). However, those approaches are still in their infancy and when applied to formalin-fixed fishes conserved in ethanol, success cases are recent (Muschick et al., 2022). An alternative option is to investigate genetic diversity using small and relatively universal barcode (mini-barcode) sequences (Shokralla et al., 2015). Shokralla and colleagues designed primers for PCR (Polymerase Chain Reaction) amplification of the mini-barcodes of the COI gene in fish. They already used these universal barcode sequences to identify species with some success. However, universal mini-barcodes might lack the required amount of polymorphisms to delineate otherwise “genetically distinct” species. On the other hand, using several different mini-barcodes might potentially solve this issue.

[1] https://github.com/amkozlov/raxml-ng
[2] https://github.com/NBISweden/MrBayes

1.4. Research questions and hypothesis

In this study, we will contribute to the detangling of the taxonomic complex diversity in *Enteromius* from West Africa. Bamba (2012) already highlighted some differences in West-African specimens that were thought to be one species. In this study, we will revise the *E. ablabes/E.* sp. 'ndoubai' and the *E. guildi/E.* sp. 'snoeksi' complexes, as found by Bamba (2012) by using an integrative approach of morphology and genetics. For the morphological approach, we will follow Bamba (2012) and Maetens et al., (2020). Genetic approaches might be crucial in exploring the diversity of *Enteromius* and ethanol-preserved specimens for the aforementioned species are available in the collections of the RMCA. In this study, we will explore the feasibility of using formalin-fixed specimens in genetic analyses for species delineation. We will i) test the feasibility of extracting DNA from formalin-fixed specimens in *Enteromius* and ii) test the amplification of mini-barcode sequences from these DNA extracts.

We will be examining specimens collected by Bamba (2012) but expanding the geographical range to cover the whole distribution of the species. Thereby, we hypothesize that we will find more morphogroups within the species complexes. Furthermore, we will use the *Enteromius* species complexes as a first case study to explore the feasibility of extracting and sequencing DNA out of formalin-fixed ethanol-preserved specimens in the RMCA. This would open new possibilities and lead to the result where it is possible to have the hundreds of thousands of formalin-fixed specimens of the RMCA collection in Tervuren available for sequence-based phylogenetic and taxonomic research.

2. Materials and methods

2.1. Study organisms

2.1.1. Enteromius ablabes

Enteromius ablabes (Bleeker, 1863) can be recognized by two pairs of moderately developed barbels, a complete lateral line, numerous well-visible cephalic pores, and a scale formula of (3.5 / 23-29 / 3.5; 2.5; 12 (11) (Teugels & Hopkins, 2007; Bamba, 2012) (Figure 2A). The scale formula is noted as follows: a / b / c; d; e with a (the number of scales between the lateral line and the origin of the dorsal fin), b (the number of scales on the lateral line), c (the number of scales between the lateral line and the belly, d (the number of scales between the lateral line and the origin of the pelvic fin), e (the number of scales around the caudal peduncle). The lateral line can be horizontal or slightly depressed below the dorsal fin (Bamba, 2012). The dorsal fin is colourless, has a slightly concave border with a smooth and flexible third unbranched dorsal fin ray, and 8 branched fin rays (Froese & Pauly, 2022). Variation in colour pattern (e.g. variation in series of black, vertical lines on the plain body) and differences in body shape (e.g. a more streamlined body shape) within and between populations was already noticed (Daget & Iltis, 1965; Lévêque, 1983; Lévêque et al., 1990). Living specimens display yellowish to orange to colourless fins and have a red colour in the eye (Bamba, 2012). The maximum size reported is 96mm in standard length (Froese & Pauly, 2022). *Enteromius spurelli* (Boulenger, 1913) was put in synonym with *Enteromius ablabes* by Lévêque in 1983. As mentioned in the introduction, *E. ablabes* has a widespread distribution in West Africa, reaching from the upper reaches of the Senegal Basin (Senegal) in the West up to the Cross River in Nigeria and Cameroon in the East (Lévêque, 2003).

2.1.2. Enteromius guildi

Enteromius guildi (Loiselle, 1973) has, like *E. ablabes*, a complete lateral line which is horizontal or slightly depressed below the dorsal fin (Froese & Pauly, 2022) (Figure 2C). The two pairs of barbels are short; the anterior pair can sometimes extend beyond the anterior margin of the eye, while the posterior pair can reach beyond the centre of the eye (Bamba, 2012). The scale formula is 3.5/ 22-25/ 3.5; 2.5; 10. The distinguished character between *E. guildi* and *E. ablabes* is the number of scales around the caudal peduncle (10 in *E. guildi* and 12 in *E. ablabes*). Like *E. ablabes, E. guildi* has a huge variation in colouration from, clear to dark, between and within populations (Bamba, 2012). The dorsal fin has a slightly concave border, is colourless and has 8 branched fin rays. The third unbranched dorsal fin ray is smooth and flexible (Froese & Pauly, 2022). The maximum size reported is 56 mm in standard length (Froese & Pauly, 2022). The species has a more restricted distribution, from an unknown

locality in Liberia and the Néro basin in Ivory Coast eastwards to the Volta basin (Ghana) (Lévêque, 2003; Bamba et al., 2011).

2.1.3. Enteromius sp. 'ndoubai'

Enteromius sp. 'ndoubai' is not yet accepted as a valid species but has been described by Bamba in 2012, who distinguished *E.* sp. 'ndoubai' from *E. ablabes*. Although this species is very similar to *E. ablabes*, it can be distinguished by the longer barbels, the absence of the cephalic sensory canals and the larger caudal peduncle (Bamba, 2012) (Figure 2B). The scale formula is 3.5/ 22-24/ 3.5; 2.5; 12 and the preserved specimens have an overall brownish ground colour (Bamba, 2012). Specimens of *E.* sp. 'ndoubai' have a complete lateral line, which is horizontal or can be slightly depressed below the dorsal fin. The third unbranched dorsal fin ray is smooth and flexible. The dorsal fin has a slightly concave border, 8 branched fin rays and no clearly defined spots (Bamba, 2012). The maximum size observed by Bamba is 79.1 mm in standard length. *Enteromius* sp. 'ndoubai' is currently only known in the Konkouré basin (Guinea) (Bamba, 2012).

2.1.4. Enteromius sp. 'snoeksi'

Like *E.* sp. 'ndoubai', *Enteromius* sp. 'snoeksi' is described by Bamba in 2012 and is not yet a valid species. Bamba (2012) distinguished *E.* sp. 'snoeksi' from *E. guildi* by a larger eye and very short barbels. *Enteromius* sp. 'snoeksi' has in general small scales, with a scale formula of 3.5/ 25-30/3.5; 2.5; 10 (11) (Bamba, 2012) (Figure 2D). The dorsal fin has a slightly concave border, is colourless and has 8 branched fin rays. The paired and unpaired fins are whitish transparent and there is no black longitudinal band on the body flank (Bamba, 2012). The maximum observed size is 56.5 mm in SL by Bamba (2012). The species is currently only known to be found in the Tanoé River in Ivory Coast (Bamba, 2012).

2.2. The geographic region of samples studied

Bamba (2012) studied the species complexes *E. ablabes/E.* sp. 'ndoubai' and *E. guildi/E.* sp. 'snoeksi', concentrating on the populations from Ivory Coast. To further explore these complexes in a wider, more balanced geographical context, I examined specimens of the four species from all over West Africa, ranging from Guinea in the West up to Cameroon in the East (Figure 3). First, all specimens of the interested species registered in the RMCA collection in Tervuren were listed and mapped in QGIS. Specimens of *E.* sp. 'ndoubai', *E. guildi* and *E.* sp. 'snoeksi' from all locations, present in the collection of the MRAC, were examined. Specimens of *E. ablabes* were selected based on their geographic distribution and standard length (> 35 mm). Up to five specimens per location were studied. No specimens of *E. ablabes* from the more Western part of the research area and only 15 specimens of *E. guildi* were registered in

the RMCA collection. Therefore, a study visit to the Museum national d'Histoire naturelle (MNHN) in Paris was arranged to examine specimens of *E. guildi* and *E. ablabes* from the regions of interest. A more detailed list of the used specimens from both collections can be found in Addendum C.

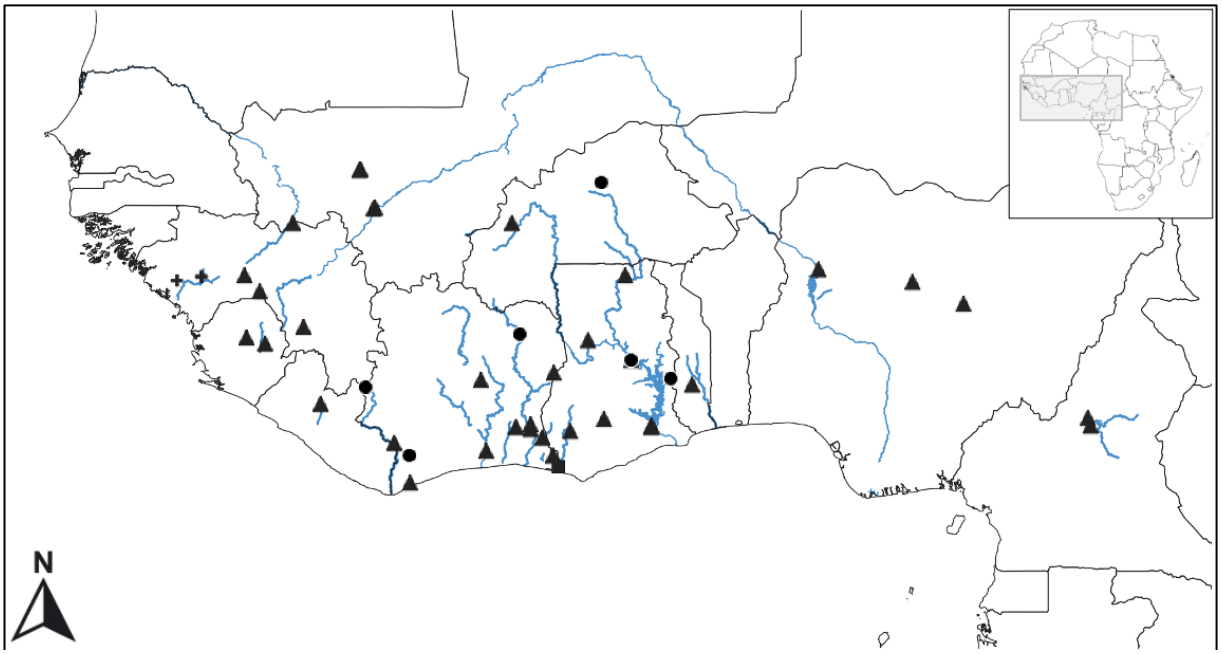

Figure 3: Map of all sampling locations in West Africa. The river systems explored in this thesis are in blue. The triangles represent the sampling locations of *E. ablabes*, the circles *E. guildi*, the crosses *E.* sp. 'ndoubai' and the squares *E.* sp. 'snoeksi'. For more information, see Addendum C with a list of all used specimens and locations. To represent the rivers, we used the shape files of DIVA-GIS.

2.3. Morphometric approach

In total, 173 specimens were examined for morphological analyses. On each specimen, 26 morphometric measurements and thirteen meristics were performed. Both the measurements and meristics were based on Maetens et al., (2020) which was inspired by Bamba et al., (2011) and Bamba (2012). When examining specimens, several states were observed. Some specimens were extremely fragile and nearly fell apart. In these specimens, the ventral side was often on the verge of rupturing. On the other hand, specimens were found that were firm. Still, the morphometrics were obtained in the bast possible way from each specimen of both states. Measurements and meristics were taken on the taxonomic, left side of the specimens. If a structure was broken or invisible, we measured it on the right side of the specimen. A detailed description of the measurements and the meristics is given below.

2.3.1. Measurements

We used a Mitutoyo calliper with a precision of 0.01 mm to take measurements of the specimens. The length of the anterior barbel and posterior barbel were measured using a

binocular and a Schott light source. The measurements listed in Table 1 and shown in Figure 4 were performed on each specimen.

Table 1: The measurements taken during this study, are based on the text of Maetens et al., (2020).

Trait	Description
Standard length (SL) (Figure 4a, nr1)	Distance from the anterior tip of the snout to the vertical line through the most posterior point of the caudal peduncle.
Head length (HL) (Figure 4a, nr2)	Distance from the anterior tip of the snout to the posterior most bony point of the operculum.
Eye diameter (ED) (Figure 4a, nr3)	The horizontal eye diameter.
Snout length (SnL) (Figure 4a, nr4)	Distance from the anterior tip of the snout to the anterior margin of the eye.
Pre-operculum distance (PrOpD) (Figure 4a, nr5)	Distance from the anterior tip of the snout to the posterior margin of the pre-operculum.
Interorbital width (IOW) (Figure 4b, nr6)	The minimum width between the orbits.
Pre-dorsal distance (PrDD) (Figure 4a, nr7)	Distance from the anterior tip of the snout to the articulation of the first dorsal-fin ray.
Post-dorsal distance II ($PoDD_2$) (Figure 4a, nr8)	Distance from the articulation of the last dorsal-fin ray to the vertical line through the most posterior point of the caudal peduncle.
Dorsal fin base length (DoFBL) (Figure 4a, nr9)	Distance between the articulation of the first dorsal-fin ray and the last dorsal-fin ray.
Dorsal fin length (DoFL) (Figure 4a, nr10)	Distance from the articulation of the first dorsal-fin ray to the proximal tip of the longest dorsal-fin ray.
Pre-pectoral distance (PrPecD) (Figure 4a, nr11)	Distance from the anterior tip of the snout to the articulation of the first pectoral-fin ray.
Pre-pelvic distance (PrPelD) (Figure 4a, nr12)	Distance from the anterior tip of the snout to the articulation of the first pelvic-fin ray.
Pre-anal distance (PrAD) (Figure 4a, nr13)	Distance from the anterior tip of the snout to the articulation of the first anal-fin ray.
Post-anal distance II ($PoAD_2$) (Figure 4a, nr14)	Distance from the articulation of the last anal-fin ray to the vertical line through the most posterior tip of the caudal peduncle.
Body depth II (BD_2) (Figure 4a, nr15)	The depth of the body at the articulation of the first dorsal-fin ray.
Maximum caudal peduncle depth (MxCPD) (Figure 4a, nr16)	The depth of the caudal peduncle at the articulation of the last anal-fin ray.
Minimum caudal peduncle depth (MnCPD) (Figure 4a, nr17)	The minimal depth of the caudal peduncle.

Pre-occipital distance (PreOcD) (Figure 4a, nr18)	Distance from the anterior tip of the snout to the posterior occipital margin.
Head width (HW) (Figure 4b, nr19)	The width of the head at the level of the posterior margin of the pre-operculum.
Anal fin base length (AnFBL) (Figure 4a, nr20)	Distance between the articulation of the first anal-fin ray and the last anal-fin ray
Anal fin length (AFL) (Figure 4a, nr21)	Distance from the articulation of the first anal-fin ray to the proximal tip of the longest anal-fin ray.
Pectoral fin length (PcFL) (Figure 4a, nr22)	Distance from the articulation of the first pectoral-fin ray to the proximal tip of the longest pectoral-fin ray.
Pelvic fin length (PlvFL) (Figure 4a, nr23)	Distance from the articulation of the first pelvic-fin ray to the proximal tip of the longest pelvic-fin ray.
Head depth (HD) (Figure 4a, nr24)	The depth of the head at the level of the occipital margin.
Length of the anterior barbel (BarbA)	Distance from the articulation of the anterior barbel to the proximal tip of the anterior barbel.
Length of the posterior barbel (BarbP)	Distance from the articulation of the posterior barbel to the proximal tip of the posterior barbel.

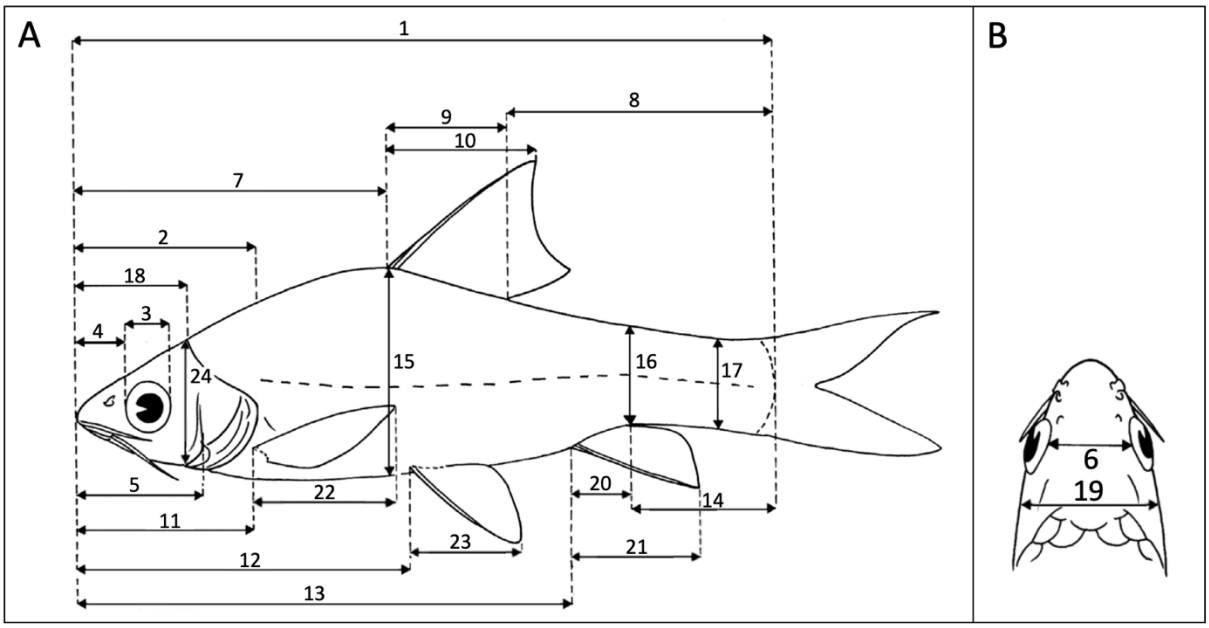

Figure 4: Schematic illustration of the morphometric measurements that were taken during this study. Illustrations edited from Bamba et al., (2011).

2.3.2. Meristics

All meristics were counted under a binocular with the assistance of a Scott light source. The meristics used in this thesis are based on Bamba (2012) and Maetens (2019) (Table 2) (Figure 5). While counting the dorsal and anal fin rays, I observed differences when counting on the base of the fin compared to counting on the middle of the fin. Both methods are being used for counting the dorsal and anal fin rays in literature (Armbruster, 2012; Englmaier et al., 2020a).

After an X-ray examination of some *E. ablabes* specimens, I decided to follow Armbruster (2012) by counting from the base of the dorsal and anal fin. Furthermore, we used an encoding system based on the relative position of the extremity of the barbels when they are strung along the head of the specimen, based on Bamba et al., (2011) (Table 3).

Table 2: Schematic table of the counts/meristics that were taken into account in this study, based on Bamba et al., (2011).

Trait	Description
Lateral line scales (LLS) (Figure 5A, LLSc)	The number of scales counted in the lateral and/or longitudinal line from the gill cover to the posterior tip of the caudal peduncle, scales after this point are indicated after with a plus sign (LLS_total)
Scales dorsal fin – lateral line (D-LSc) (Figure 5A, D-L Sc)	The number of scales counted from the articulation of the first dorsal-fin ray to the lateral line following a posterior-ventrally series down, the scale at the articulation of the first dorsal-fin ray is counted as 0.5 and the scale on the lateral line is not counted.
Pre-dorsal scales (PDSc) (Figure 5A, PD Sc)	The number of scales counted from the occiput to the articulation of the first dorsal-fin ray, the scale at the articulation of the first dorsal-fin ray in included in the counting.
Scales around caudal peduncle (CPS) (Figure 5A, CP Sc)	The number of scales counted in a transversal series encircling the caudal peduncle at the level of the minimal caudal peduncle depth.
Scales lateral line – pelvic fin (L-PSc) (Figure 5A, L-P Sc)	The number of scales counted from the lateral line to the articulation of the first pelvic-fin following an antero-ventrally series down, the scale on the lateral line is not included and the scale on the articulation of the pelvic-fin is included.
Scales lateral line - belly (L-BSc) (Figure 5A, L-B Sc)	The number of scales counted from the lateral line to the middle of the belly following an antero-ventrally series down, the lateral line scale is not included in the counting and the most ventral scale is counted as 0.5.
Dorsal fin rays (DFR)	The number of rays counted at the bases of the dorsal fin, with the unbranched fin rays in Roman numerals and the branched fin rays in Arabic numerals.
Anal fin rays (AFR)	The number of rays counted at the bases of the anal fin, with the unbranched fin rays in Roman numerals and the branched fin rays in Arabic numerals.
Pectoral fin rays (PecFR)	The number of rays counted at the middle of the pectoral fin, with the unbranched fin rays in Roman numerals and the branched fin rays in Arabic numerals.
Pelvic fin rays (PelFR)	The number of rays counted at the middle of the pelvic fin, with the unbranched fin rays in Roman numerals and the branched fin rays in Arabic numerals.

Gill rakers on upper limb of first gill arch (GrUp)	The number of gill rakers counted on the upper limb of the first gill arch; the gill raker placed on the articulation between the upper and lower limb is not included.
Gill rakers on lower limb of first gill arch (GrLow)	The number of gill rakers counted on the lower limb of the first gill arch; the gill raker placed on the articulation between the upper and lower limb is not included.
Total number of gill rakers (GrTot)	The total number of gill rakers counted on the first gill arch, including the gill raker placed on the articulation between the upper and lower limb.
Encoding of the anterior barbel (code_BarbA) (Figure 5B)	Encoding system based on the extremity of the anterior barbel when stretch along the head, see Table 3
Encoding of the posterior barbel (code_BarbP) (Figure 5B)	Encoding system based on the extremity of the posterior barbel when stretch along the head, see Table 3

Table 3: Encoding systems of the barbels based on Bamba et al., (2011).

Code number (Figure 5B)	Description
Code 1	Code 1 is given to the part of the head anterior to the ventral line through the nostril.
Code 2	Code 2 is given to the part between the nostril and the vertical line through the anterior margin of the eye.
Code 3	Code 3 is given to the part between the vertical line through the anterior margin of the eye and the vertical line through the eye centre.
Code 4	Code 4 is given to the part between the vertical line though the eye centre and the vertical line through the posterior margin of the eye.
Code 5	Code 5 is given to the part between the vertical line through the posterior margin of the eye and the vertical line through the posterior margin of the pre-operculum.
Code 6	Code 6 is given to the part between the vertical line through the posterior margin of the pre-operculum and the vertical line through the centre of the operculum.
Code 7	Code 7 is given to the part between the vertical line through the centre of the operculum and the vertical line through the posterior margin of the operculum.
Code 8	Code 8 is given to the part posterior to the vertical line through the posterior margin of the operculum.

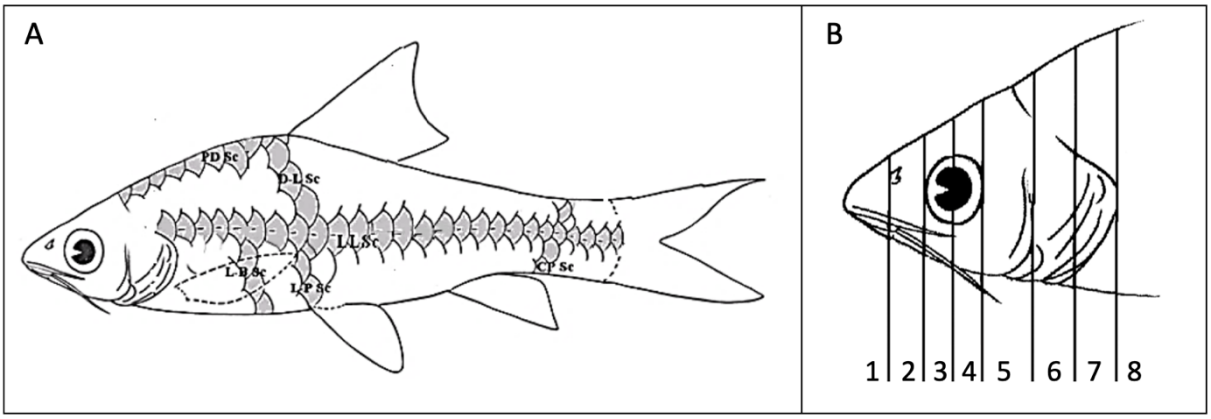

Figure 5: Schematic illustrations of the meristics that were taken during this study. Figure 5A edited from Bamba (2012) and Figure 5B edited from Bamba et al., (2011).

2.3.3. Data analyses

Multivariate data analyses

A Principal Component Analysis (PCA) was used to explore the multivariate dataset. A large number of variables was reduced by the PCA into meaningful principal components, which are a linear combination of the variables (Manly, 1994; Snoeks, 2004). The largest variation is explained by the first principal component, and the explanation of variation decreases when further principal components are added (Manly, 1994). The PCA was performed separately on measurements and meristics. For the analysis of the measurements, both the log-transformed data and the data expressed as percentages of standard length (% SL) were used. Unless the % SL data better highlighted a certain finding, the result of the PCA using the log-transformed data is presented in the results. For analysing the measurements, the covariance matrix was used. We utilized the raw data for the meristics, using the correlation matrix. For the log-transformed measurements, the first principal component was not included as it can be interpreted as a proxy for size (Bookstein et al., 1985). The measurements of the barbels were excluded from the analyses as these values had too much influence in the PCA on log-transformed data. Barbels are known to break off easily, still with the use of the codes for the barbels introduced by Bamba et al., (2011), these structures are taken into account. Furthermore, as some meristics were linked with each other, we decided to keep the LLS_total and the GrUp and GrLow, while excluding the LLS and the GrTot. The constant values in the analyses of the meristics were also excluded from the PCA. All loadings of the PCs with an absolute value higher than 0.3 were listed in the results. Only the six highest loadings of both PCs combined were shown in the biplots of the PCAs.

Univariate data analyses

To compare variables between different groups, we performed non-parametric Mann-Whitney U (MWU) tests with a sequential Bonferroni correction according to Rice (1989). We examined

which variables were significantly different between *E. ablabes* and *E. guildi*, between *E. ablabes* and *E.* sp. 'ndoubai', and between *E. guildi* and *E.* sp. 'snoeksi'. We first ensured that the dataset was balanced for each comparison, meaning the p-value of the MWU test of the SL was close to 0.5. To create these datasets, larger or smaller specimens were randomly removed from the datasets until this threshold was reached. After sequential Bonferroni correction, p-values lower than 0.05 were seen as significant and p-values lower than 0.01 were seen as highly significant. Furthermore, individual scatterplots of the variables in the function of the standard length (SL) were made to visualize and compare the results of the MWU tests. The data used for the scatterplots is the data of the measurements expressed as percentages of SL and the raw data for the meristics. A range of 35 mm to 65 mm in SL was set in the scatterplots to compare the specimens with the same SL. Thereby, interfering outliners were excluded from the visual comparison in the scatterplots.

The statistical analyses were performed in RStudio version 2021.09.0-351 with R version 4.1.1 (2021- 08-10). In the R software, the following packages were used: 'base', 'cluster', 'dplyr', 'factoextra', 'FSA', 'ggplot2', 'ggpubr', 'ggthemes', 'lattice' and 'vegan'.

2.4. Molecular approach / DNA barcoding

2.4.1. DNA extraction on formalin-fixed ethanol-preserved specimens.

Sample selection and working with batches

From the 173 specimens measured for the morphometric approach, we selected 60 specimens to perform DNA extraction. The selection was based on sampling location, in order to get a balanced and representative dataset on the whole distribution range. To take into account the genetic variation at the local scale, two to three specimens were used from each selected sampling location. The amount of tissue needed for DNA extraction is relatively large (i.e. cut-out of 1 cm on 1 cm of muscle tissue), which results in considerable damage in small specimens. Therefore, only specimens with a SL of 40 mm or higher were chosen.

We carried out the DNA extraction in four batches (Table 4). This allowed us to work with fewer samples at a time, minimizing and preventing contamination from spreading widely among extracts. For each batch, we selected samples covering the whole geographic range in West Africa to remove the batch effect. It allowed us to make sure that if we detected a geographical structure in the data, it is not due to cross-contamination within the same batch of extraction. The number of specimens per batch varied. The first batch included a limited number of samples (n=6) and allowed us to optimize the protocol and build up experience on the extraction kit (see further). The samples of *E. guildi* and *E.* sp. 'snoeksi' were included in batches 3 and 4 due to the limited number of specimens available in the RMCA. A negative

control was included in each batch (c-B#). Information on the dates and species used in the different batches is provided in Table 4

Table 4: Detailed overview of the different batches performed for the DNA extraction.

	Batch 1	Batch 2	Batch 3	Batch 4
Date (starting-end)	29/11/2022 - 02/12/2022	19/12/2022 - 22/12/2022	31/01/2023 - 03/02/2023	20/02/2023 - 23/02/2023
Number of *E. ablabes* specimens	6	10	9	19
Number of *E.* sp. 'ndoubai' specimens	0	0	3	4
Number of *E. guildi* specimens	0	0	3	4
Number of *E.* sp. 'snoeksi' specimens	0	0	0	2
Total number of specimens	6	10	15	29
Scales or skin removed	NO	YES	YES	YES

From batch 2 onwards, we removed the scales and/or skin present optimise the working of the Proteinase K.

DNA extraction

The RMCA in Tervuren does not have a dedicated laboratory for working on ancient/degraded DNA, and we minimalised the likelihood of contamination as much as possible. Before the extraction, all equipment and consumables were exposed to UV-C light (10-30 minutes). The benches were cleaned with bleach (5-10 minutes) and 70% ethanol to remove potential contaminants. Furthermore, systematically, filter tips were used and each dissection was conducted using a new and flame-sterilised razor blade. Tweezers were also systematically flame-sterilised and cleaned between dissections. We performed the extractions under a UV-C hood and gloves were worn throughout the whole extraction process and were regularly changed.

The protocol performed in this thesis is a modified version of Diedericks (2017) based on the QIAamp® DNA FFPE Tissue Kit (QIAGEN). A sample of 1 cm on 1 cm of muscle tissue below the dorsal fin on the right side of the fish was taken from each specimen. As recommended by Pikor et al., (2011) and Diedericks, (2017), rehydration of the tissue was first conducted using several steps of ethanol and ultra-pure sterile (UPS) water baths. The rehydration steps were executed with a first bath of 100% ethanol, followed by a bath of 70% ethanol and by UPS water bath. As a last step of the rehydration, the muscle tissues were soaked in UPS water at 55°C for two days in contrast to the five days in Diedericks (2017). Furthermore, the amount of Proteinase K specified in the QIAGEN protocol was used rather than the double amount as proposed by Diedericks (2017). Elution was performed using 20 µL of elution buffer dispensed onto the centre of the column membrane and left to incubate at room temperature for five minutes before centrifuging for two minutes at full speed (13300 rpm). Elution was conducted two times for a total volume of 40 µL DNA. The concentration of DNA was quantified for each

extract using Qubit 1X dsDNA High Sensitivity (HS) assay (Invitrogen). Measurements were conducted on a Qubit 4 Fluorometer using 2 µL of DNA extract each time. All DNA extractions were stored at -20°C. For a more thorough description of the DNA extraction protocol used, see Addendum D.

Ruane and Austin, (2017) were able to sequence DNA samples from formalin-fixed specimens with a DNA concentration above 1 ng/µL. Based on these findings, we classified our DNA extracts as the following: extracts with a DNA concentration below 0.550 ng/µL were labelled "Very Low", extracts with a DNA concentration between 0.550 and 1.000 were labelled "Low", extracts with a DNA concentration between 1.000 and 2.000 were labelled as "Okay", and extracts with a DNA concentration higher than 2.000 were labelled as "High".

2.4.2. Sequencing of the extracted DNA

Choice of primers for mini-barcoding: Mini-D and Mini-E

With a view to future genetic research on formalin-fixed fish specimens of the RMCA collection, we decided to investigate and test universal primers for a wide range of fish taxa rather than primer pairs for a specific species. The universal primers used in this work are effective on a vast range of fish, meaning they could also be effective for other groups than *Enteromius* in future research. Additionally, the universal primers are reported to be robust for degraded DNA (Shokralla et al., (2015), but see Appleyard et al., (2021)). We tested two sets of primer pairs of the ones developed by Shokralla et al., (2015): Mini-D and Mini-E. The selection of the two primer pairs was based on the fragments of the CO1 gene they amplify and the amplification/sequencing success rates in Shokralla et al., (2015). Both primers target regions that do not overlap, leading to better coverage of the COI gene (Figure 6). The primers are designed to be M13-tailed, meaning it allowed us to get more of the sequence of interest during Sanger sequencing (Shokralla et al., 2015). Additionally, with the use of the M13-tails, we only needed one pair of primes for sequencing, while targeting different barcodes. According to the findings of Shokralla et al., (2015), the Mini-D and Mini-E primer pairs appeared to operate more efficiently to discover polymorphism than the other designed primer pairs. The expected length of each mini-barcode was around 200 bp as it corresponds to the size of DNA fragments retrieved in museum specimens (Englmaier et al., 2020b). An overview of the Mini-D and Mini-E primer pairs is provided in Table 5.

Table 5: Names, sequences and annealing temperature of the primers (forward and reverse) used in this thesis (from Shokralla et al., (2015)).

Primer Set	Primer name	Direction	Primer sequence (5'-3')	Annealing temp. (°C)
Mini-D	Mini-D_F	Forward	CACGACGTTGTAAAACGACGGIACIGG ITGRACIGTITAYCCYCC	50

	Mini-D_R	Reverse	GGATAACAATTTCACACAGGGTRATIC CIGCIGCIAGIAC	
Mini-E	Mini-E_F	Forward	CACGACGTTGTAAAACGACACYAAICA YAAAGAYATIGGCAC	46
	Mini-E_R	Reverse	GGATAACAATTTCACACAGGCTTATRT TRTTTATICGIGGRAAIGC	

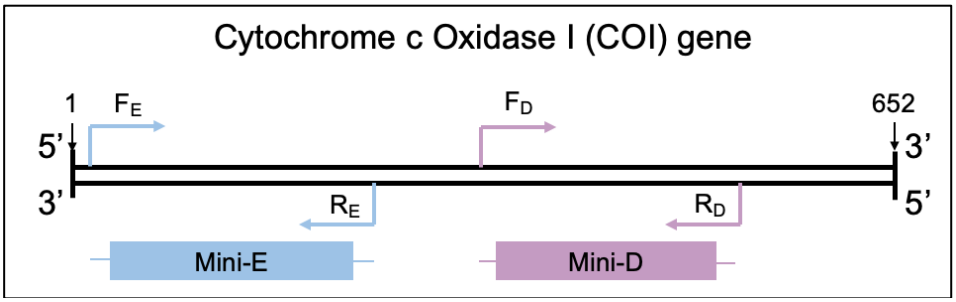

Figure 6: Schematic representation of the regions in the Cytochrome c Oxidase I (COI) gene (652bp) amplified by the mini barcodes: Mini-E and Mini-D. F = forward primer; R = reverse primer. Figure adapted from Shokralla et al., (2015).

PCR amplification: touchdown and "two-time" PCR

The PCR (Polymerase Chain Reaction) amplification of the mini-barcodes can be challenging due to the nature of the DNA templates at hand (low concentration of degraded DNA). To get PCR products that are suitable for Sanger sequencing, we tested two different approaches. We tested a touchdown PCR, a PCR method using first a small number of cycles with "high" temperature during the annealing step, to amplify more stricly specific sequences, then most cycles are conducted with lower annealing temperature. The touchdown PCR was tested to improve the specificity of the primers' action while requiring a smaller amount of primers and enzymes. This method has already proved useful for cases involving hDNA (historical DNA) (Englmaier et al., 2020b). We followed the same PCR programme as Englmaier et al., (2020b) with slight modifications regarding the temperature profiles for the annealing steps (see Shokralla et al., (2015)). As an alternative method, a "two-time" PCR was tested. First, we applied a classic PCR programme to generate a first set of amplified DNA. The PCR product resulting from this PCR was then used for a second round of PCR (same programme). Even if this approach is prone to many issues (background amplification, high level of PCR errors (etc.), it proved to be of some use for cases where DNA is highly degraded with low concentrations (Diedericks, pers. commun.). Another downside of this "two-time" PCR method is the cost, as it consumes twice the amount of reagent than the touchdown PCR. The PCR programme was also based on Englmaier et al., (2020b) with slight modifications according to Shokralla et al., (2015) (Table 6). The PCR programmes for both methods and both primer pairs are detailed in Table 6.

A positive control and two different kinds of negative controls were included in both PCR methods. A DNA extract already successfully used for cyt b and CO1 barcoding in other projects was included to control that the PCR programmes properly amplify the targeted mini-barcodes (positive control). As the primers are universal for fishes and as we did not have any *Enteromius* DNA extraction from a close relative to *E. ablabes* available, we did not select any particular species and took what was available in the lab (*Hydrocynus forskahlii* (Cuvier, 1819) from the HIPE collection (HP4048)). For the negative control, we included the negative control from the DNA extraction step (c-B#), plus a "PCR" negative control (C-), corresponding to the PCR mix with some UPS water instead of DNA extraction. The different PCRs were conducted on a Biometra Tone Thermal Cycler. The details of the PCR mix composition are presented in Table 7 and are based on Englmaier et al., (2020b). We used only half of the reagent amount per reaction mentioned by Englmaier et al., (2020b) to decrease the cost, as the AmplitaqGold360 PCR reagents are relatively expensive. The different PCR products were controlled and checked on agarose gel (1%) electrophoresis.

As we had a limited amount of PCR reagents allocated to our exploratory experiments, we did not use all samples for each PCR programme. We performed more touchdown PCRs as the "two-time" PCR method can be more problematic as discussed earlier. Furthermore, we conducted PCRs on more samples using the Mini-E primer pairs than the Mini-D primer pair, due to the higher success rate of amplification and sequencing for that mini barcode, according to Shokralla et al., (2015). For the "two-time" PCR using Mini-D, six DNA extracts, a negative control and a positive control were amplified. The same samples were used for the "two-time" PCR using the Mini-E primers. All 60 extracted DNA samples were amplified with the touchdown method using the Mini-E primer, while seventeen DNA samples were amplified with the touchdown method using the Mini-D primer. Again, a PCR-negative control (UPS water) and a positive control (HP4048) were used for the touchdown method. Furthermore, for the Mini-E touchdown PCR, all four negative controls of the DNA extraction were included, while the Mini-D touchdown PCR had the negative control of the fourth extraction batch (c-B4) included. In total, 89 amplifications, excluding the negative and positive controls, were conducted.

Table 6: PCR programmes used for the "two-time" PCR (programmes 1 and 2) and touchdown PCR (programmes 3 and 4) for primer pair Mini-D and primer pair Mini-E. The programmes are based on Shokralla et al., (2015).

PCR programme 1: PCR (Mini-D)		
Number of cycles	**Temperature (°C)**	**Time**
	95	10 min
35x	95	45 sec
	50	45 sec
	72	45 sec

	72	7 min
	12	Hold ∞
PCR programme 2: PCR (Mini-E)		
Number of cycles	**Temperature (°C)**	**Time**
	95	10 min
35x	95	45 sec
	46	45 sec
	72	45 sec
	72	7 min
	12	Hold ∞
PCR programme 3: Touchdown PCR (Mini-D)		
Number of cycles	**Temperature (°C)**	**Time**
	95	10 min
5x	95	30 sec
	51	2 min
	72	45 sec
40x	95	45 sec
	50	45sec
	72	45sec
	72	7 min
	12	Hold ∞
PCR programme 4: Touchdown PCR (Mini-E)		
Number of cycles	**Temperature (°C)**	**Time**
	95	10 min
5x	95	30 sec
	47	2 min
	72	45 sec
40x	95	45 sec
	46	45sec
	72	45sec
	72	7 min
	12	Hold ∞

Table 7: The PCR mix composition for one PCR sample, based on Englmaier et al., (2020b).

Reagents*	**Amount of reagents per reaction (µL)**
AmplitaqGold360 buffer 10x	2.5
25mM Magnesium Chloride (25mM)	2
360 GC Enhancer	1
dNTP (dATP, dCTP, dGTP, dTTP, Lithium salts) (20 mM)	0.75
Primer Forward (50 µM)	0.25
Primer Reverse (50 µM)	0.25
DNA	3

AmpliTaq Gold 360 DNA Polymerase	0.2
UPS Water	15.05
Total	25

(*) Reagents from the AmplitaqGold360 “kit”. Concentration provided in the first column corresponding to the stock solution concentration. The primers used in the PCR mix depend on the mini-barcode targeted and are detailed in Table 5.

Sanger sequencing, sequences edition and blasting

As mentioned earlier, this work is exploratory and we only conducted Sanger sequencing on a limited amount of samples. Therefore, only 48 PCR products underwent bidirectional sequencing. The selection of samples for sequencing was made based on paired sampling sequencing and the outcome of the gel electrophoreses. All negative controls, expressing a band or not, were also selected. Before sequencing the amplified DNA, the PCR products were purified enzymatically using an ExoSap (Fermentes). Each chosen sample was sequenced both forward and reverse using M13 primers (M13-F: CACGACGTTGTAAAACGAC; M13-R: GGATAACAATTTCACACAGG) (Shokralla et al., 2015). The purified PCR products were then sent for Sanger sequencing to a specialised commercial company (Macrogen Europe, https://dna.macrogen.com).

The raw sequences (i.e. electropherograms) were first checked visually for quality using the program 4Peaks (https://nucleobytes.com/4peaks/index.html). When the electropherogram was difficult to read or unreadable, meaning that a lot of ambiguous base calls were present, the sequence was not further edited and was labelled as “BAD” (Figure 7B-E). Sequences assigned as such were not corrected or trimmed. If the electropherogram was readable, the corresponding sequence was labelled as “OK” (Figure 7A). The “OK” sequences were checked and were edited in case of bad calls or when there were ambiguities in the calls. Ambiguous signals were coded following the IUPAC nucleotide code (e.g. R when there is an ambiguity between A and G). If both forward and reverse sequences were classified as “OK”, they were assembled using Geneious Prime. Mismatch base calls in the assemblies were again corrected. The primers sequences were then trimmed off for each “OK” sequences (individual reverse and forward sequences or consensus sequences).

Before any edition, the raw sequences were blasted using the Nucleotide BLAST tool in GenBank (24th May 2023). This step provided two insights: i) to know if the sequences matched the targeted genus (*Enteromius*), and ii) to have a global overview of the quality of the dataset. In the Nucleotide BLAST tool, we used the default settings and the program ‘megablast’. Additionally, the corrected and trimmed sequences of the group labelled as “OK” were also blasted using the same specifications.

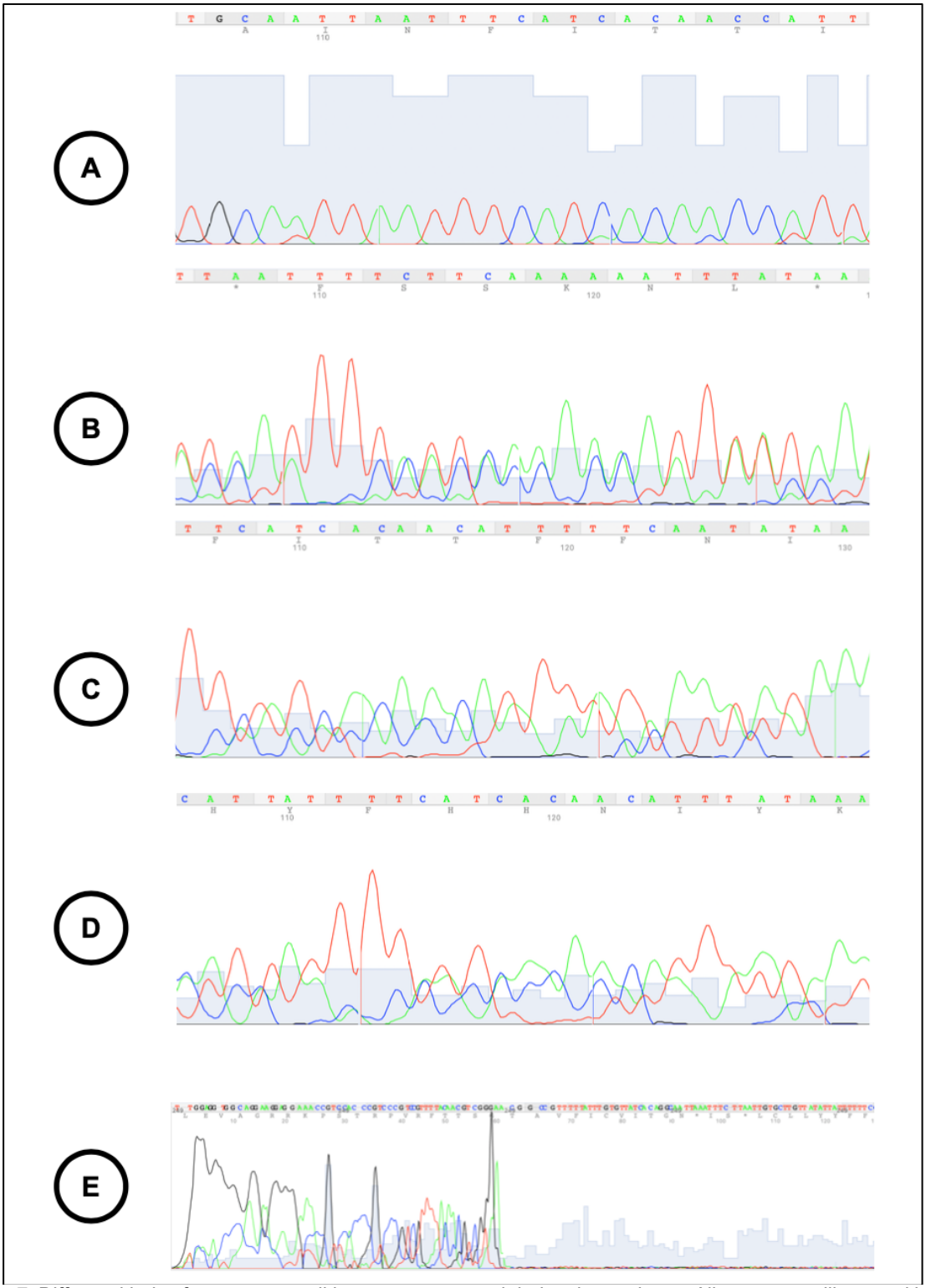

Figure 7: Different kinds of sequence qualities we encountered during the analyses. All sequences illustrated in this figure are the result of samples that underwent PCR with the Mini-D primer pair. For A) to D) electropherograms, a frame between 105 bp and 130 bp is shown. A) Positive control (forward): a good, readable electropherogram which we labelled “OK”. B) Sample DB77 (forward): An overlap in peaks in present, resulting in too much ambiguous base calls making it impossible to correct, labelled as “BAD”. C) Sample DB15 (forward): “BAD” labelled sequence as there is overlap and shifts in the traces. D) Sample DB45 (forward): even more unreadable as Figure 7C. E) PCR negative control (reverse): dimer formation visible, making it unreadable and base calls of low quality observed.

3. Results

3.1. Morphometric results

3.1.1. PCAs on data of all examined specimens

A first PCA was performed on all examined specimens of *E. ablabes, E. guildi, E.* sp. 'ndoubai' and *E.* sp. 'snoeksi' (n=173) on 23 log-transformed measurements, showing PC2 (1.9%) and PC3 (1.4%) (Figure 8). Visually, *E. ablabes* and *E. guildi* could not be distinguished in the result of this PCA. Except for one specimen of *E. guildi*, the morphospace of both species completely overlapped. The morphospace of *E.* sp. 'ndoubai' had a 50% overlap with the morphospace of *E. ablabes*. A distinction between *E. guildi* and *E.* sp. 'snoeksi' was visible (Figure 8). The highest loadings for PC2, in decreasing order, are MnCPD, DoFL and MxCPD. The highest loadings for PC3, in decreasing order, are PoAD2, ED, PoDD2, AnFBL. A PCA with the measurements (% SL) with a division based on the degrees of longitude of the sampling locations of the specimens, showed a more or less horseshoe-shaped geographical distribution on PC1 (27.6%) and PC2 (21.4%) (Figure 9). On PC3, no separation between groups was found (not illustrated). The highest loadings on PC1, in decreasing order, are PrAD, PoDD and PrDD. The highest loadings on PC2, in decreasing order, are DoFL, PlvFL and BD2. This geographic distribution was more prominent in the PCA with the % SL data (Figure 9) than in the PCA with the log-transformed data (Addendum E).

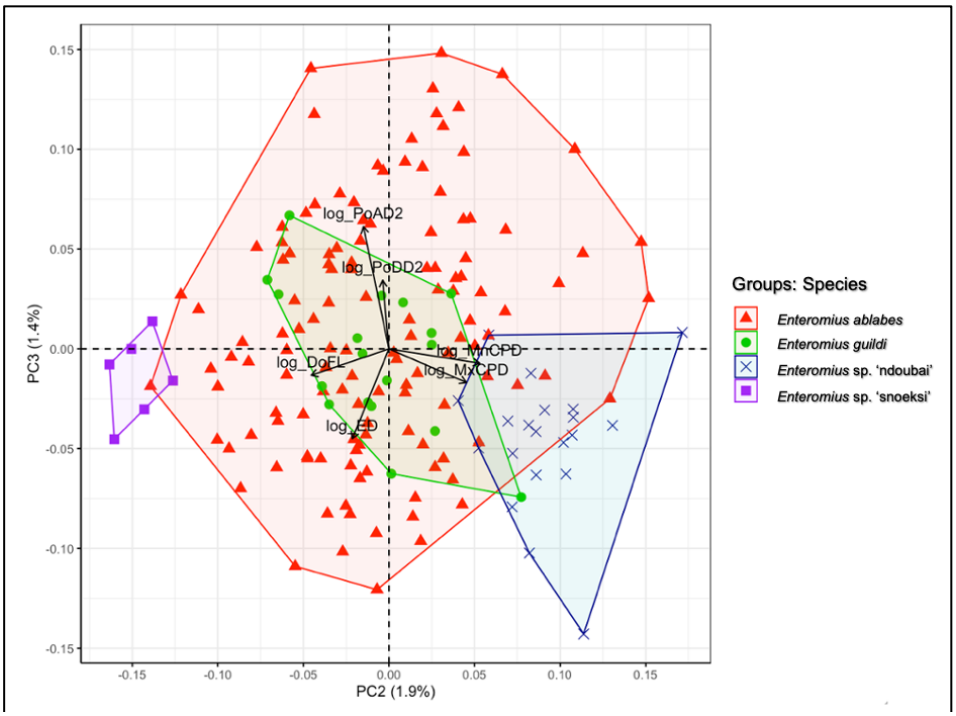

Figure 8: Biplot of a PCA on the log-transformed measurements of 129 specimens of *E. ablabes* (▲), 18 specimens of *E. guildi* (●), 20 specimens of *E.* sp. 'ndoubai' (×), 6 specimens of *E.* sp. 'snoeksi' (■). The six most important loadings of PC2 and PC3 are indicated with arrows.

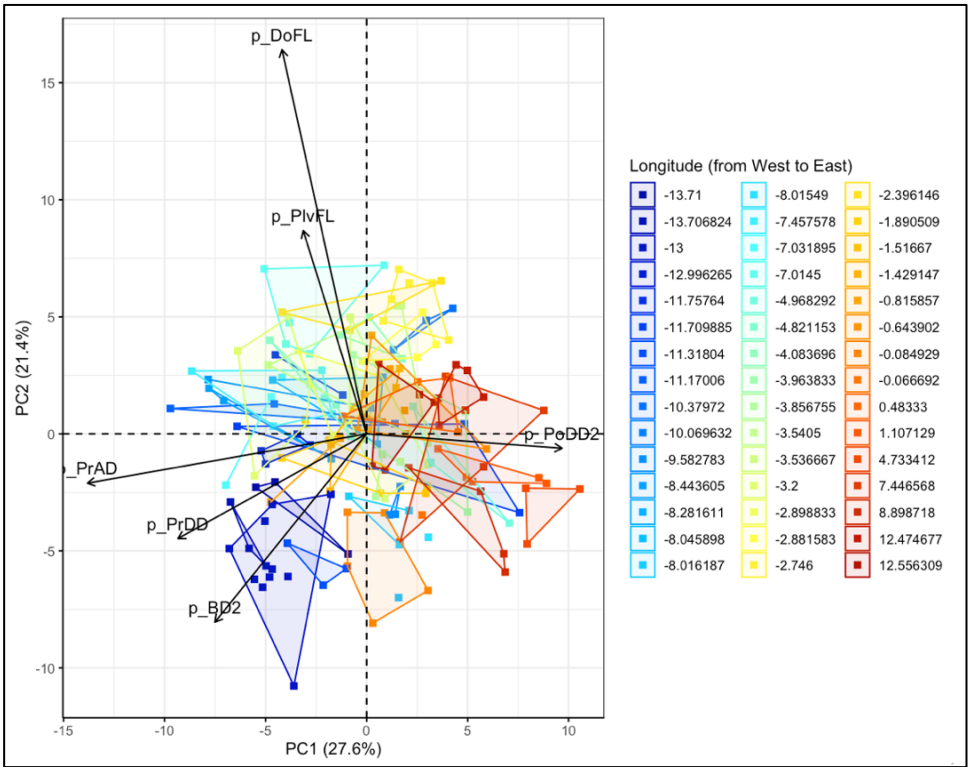

Figure 9: Biplot of a PCA on the measurements (% SL) of *E. ablabes* (n=129), *E. guildi* (n=18), *E.* sp. 'ndoubai' (n=20) and *E.* sp. 'snoeksi' (n=6) labelled with the corresponding degrees of longitude of the sampling location. The six most important loadings of PC1 and PC2 are indicated with arrows.

The result from a PCA on the meristics of all the examined species is shown in Figure 10. No distinction could be made between *E. ablabes* and *E. guildi* and between *E. ablabes* and *E.* sp. 'ndoubai'. Additionally, half the specimens of *E.* sp. 'snoeksi' were situated in the morphospace of *E. guildi*. PC1 explains 15.2% of the variation and the highest loadings, in decreasing order, are code barbel anterior, code barbel posterior and CPS. PC2 explains 11.5% of the variance and the highest loadings, in decreasing order, are LLS_total, PDSc, D.LSc, GrLow and PecFR_branched. When the groups were arranged by degrees of longitude, no geographical pattern in the biplot could be found (Figure 11).

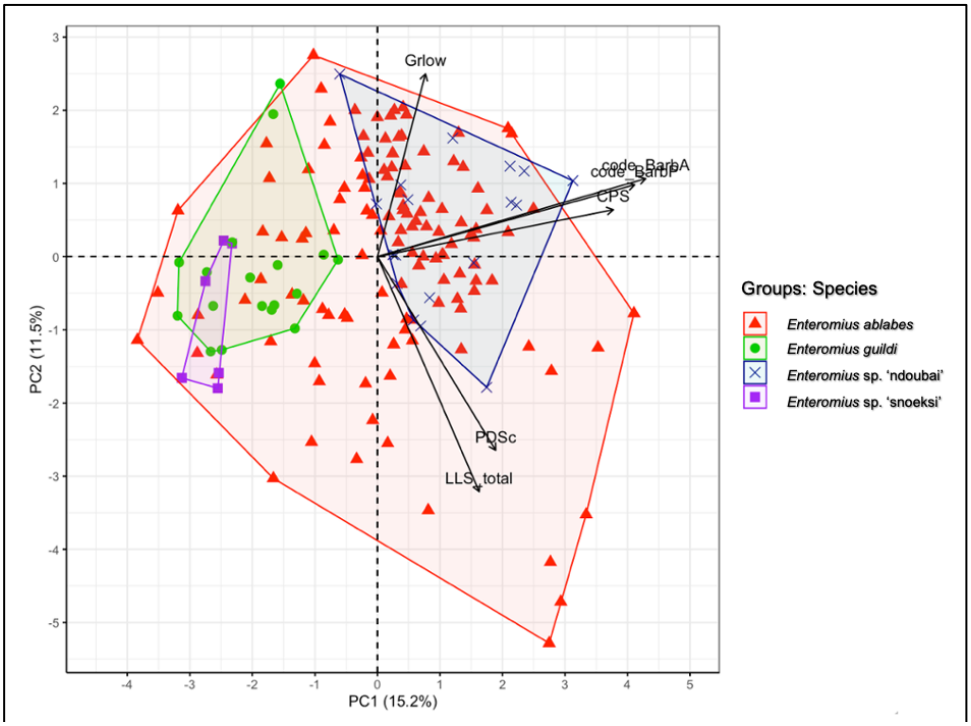

Figure 10: Biplot of a PCA on the meristics of 129 specimens of *E. ablabes* (▲), 18 specimens of *E. guildi* (●), 20 specimens of *E.* sp. 'ndoubai' (×), 6 specimens of *E.* sp. 'snoeksi' (■). The six most important loadings of PC1 and PC2 are indicated with arrows.

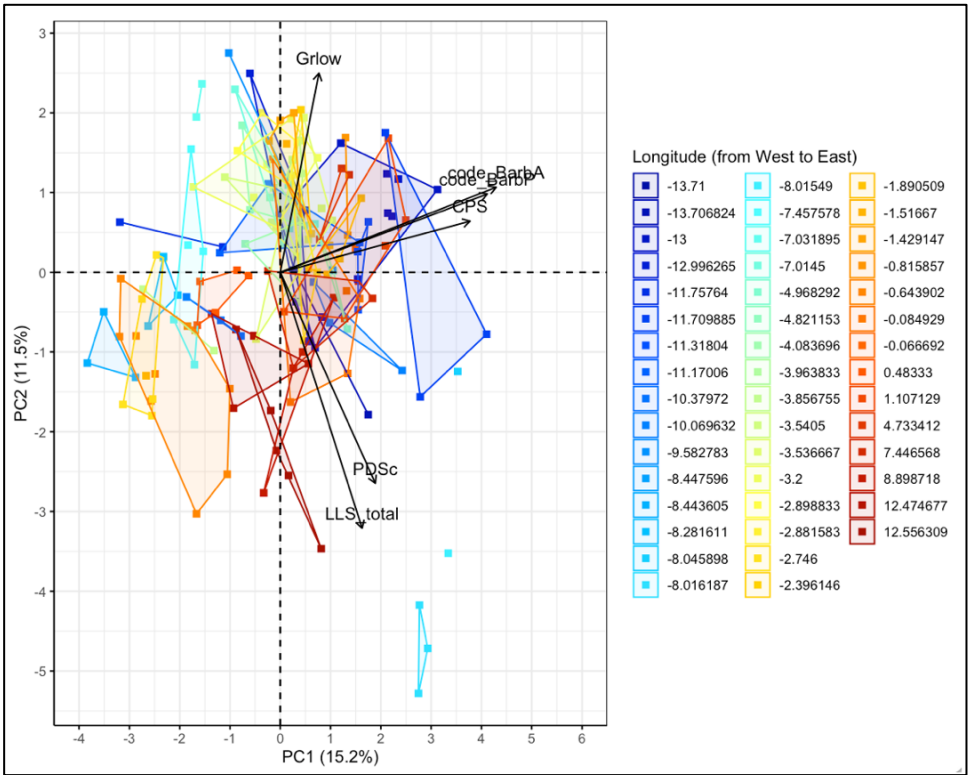

Figure 11: Biplot of a PCA on the meristics of *E. ablabes* (n=129), *E. guildi* (n=18), *E.* sp. 'ndoubai' (n=20) and *E.* sp. 'snoeksi' (n=6) labelled with the corresponding degrees of longitude of the sampling location. The six most important loadings of PC1 and PC2 are indicated with arrows.

3.1.2. Comparison between species

The results of the Mann-Whitney U-tests (MWU) with Bonferroni correction on a selection of similar-sized specimens are given in Table 8. Between *E. ablabes* (n=98) and *E. guildi* (n=18), there were no significant differences in measurements and three significant differences ($p<0.05$) in the meristics with one being highly significant ($p<0.001$). Between *E. ablabes* (n=85) and *E.* sp. 'ndoubai' (n=20), there were thirteen significant differences in total with six being highly significant. There were five significant differences in measurements and none in the meristics between *E. guildi* (n=8) and *E.* sp. 'snoeksi' (n=6).

Table 8: Results of the MWU tests with Bonferroni correction on the measurements, expressed by the percentage against SL (% SL), and the meristics between *E. ablabes* and *E. guildi*; between *E. ablabes* and *E.* sp. 'ndoubai'; and between *E. guildi* and *E.* sp. 'snoeksi'.

	***E. ablabes* (n=98) x *E. guildi* (n=18)**	***E. ablabes* (n=85) x *E.* sp. 'ndoubai' (n=20)**	***E. guildi* (n=8) x *E.* sp. 'snoeksi' (n=6)**
Measurements (% SL)			
p_HL	0.9301	0.5432	0.0293
p_ED	0.006411	0.05567	0.01998
p_SnL	0.3826	0.5762	**0.000666***
p_PrOpD	0.06782	0.5324	0.05927
p_IOW	0.2097	0.2724	**0.000666***
p_PrDD	0.02377	**6.47E-11*****	0.1079
p_PoDD2	0.5243	**0.001285***	0.2284
p_DoFBL	0.1381	0.03707	0.01998
p_DoFL	0.3234	0.005885	**0.000666***
p_PrPecD	0.5804	0.6273	0.1812
p_PrPelD	0.3744	0.09355	0.1419
p_PrAD	0.5909	**0.002567***	0.345
p_PoAD2	0.3662	**0.0006187***	0.345
p_BD2	0.9787	**3.78E-07*****	0.05927
p_MxCPD	0.7229	**7.22E-11*****	0.04262
p_MnCPD	0.3087	**1.11E-10*****	0.0293
p_PreOcD	0.06121	0.1106	0.1079
p_HW	0.9848	0.2869	**0.000666***
p_AnFBL	0.002317	0.9252	0.007992
p_AFL	0.05913	0.4358	0.01265
p_PcFL	0.1136	0.4215	0.01998
p_PlvFL	0.2875	0.1617	**0.000666***
p_HD	0.3015	**2.33E-08*****	0.004662
Meristics			

code_BarbA	**0.0005589***	**0.0008394***	0.05842
code_BarbP	**0.002052***	0.04176	0.01868
LLS_total	0.005136	**1.69E-07*****	0.3897
D-LSc	NA	0.9253	NA
PDSc	0.0575	0.01659	0.2217
CPS	**1.06E-15*****	0.242	0.1619
LPSc	0.6857	0.6542	NA
L-BSc	0.4624	0.4165	NA
DFR_unbranched	NA	NA	NA
DFR_branched	0.5568	0.1377	NA
AFR_unbranched	NA	NA	NA
AFR_branched	0.1065	**0.000129***	NA
PecFR_unbranched	0.2479	0.07186	NA
PecFR_branched	0.03323	**0.0003946***	0.09096
PelFR_unbranched	0.7112	0.4838	0.1111
PelFR_branched	0.3142	0.5331	0.3123
GrUp	0.4093	**0.003509***	0.3935
GrLow	0.04363	0.4165	0.9464

Only specimens with a SL between 35 mm and 65 mm were taken into account. Significant values ($p < 0.05$ after Bonferroni correction) are in bold and indicated with a *; highly significant values ($p < 0.001$ after Bonferroni correction) are in bold and indicated with ***. Abbreviations of the variables can be found in tables 1, 2 and 3.

Based on the MWU tests, six important traits to make a distinction between the different species were visualized using plots against SL using scatterplots (Figure 12). In Figure 12a, where HD (head depth) is plotted against SL, a distinction between *E. ablabes* and *E*. sp. 'ndoubai' was observed, with *E.* sp. 'ndoubai' having a higher HD than *E. ablabes*. Furthermore, a distinction between *E. guildi* and *E.* sp. 'snoeksi' was found, with *E.* sp. 'snoeksi' having a smaller HD than *E. guildi*. *Enteromius ablabes* and *E. guildi* could not be distinguished, based on this plot. *Enteromius* sp. 'ndoubai' had a longer PrDD than *E. ablabes* (Figure 12b). A comparison of PlvFL showed a longer PlvFL in *E.* sp. 'snoeksi' than in *E. guildi* (Figure 12c). Specimens of *E.* sp. 'ndoubai' and *E. guildi* had a higher MxCPD than the specimens of *E. ablabes* and *E.* sp. 'snoeksi' respectively (Figure 12d). A less clear distinction between species was observed when comparing the MnCPD. However, *E.* sp. 'snoeksi' had a higher MnCPD than *E. guildi* and *E.* sp. 'ndoubai' had a higher MnCPD than *E. ablabes* (Figure 12e). Figure 12f compares the number of CPS between species, where a clear distinction was made between *E. ablabes* and *E. guildi,* with specimens of *E. ablabes* having more scales around the caudal peduncle than *E. guildi.* Some specimens of *E. ablabes* (n=4) had 10 caudal peduncle scales while other specimens (from Kalaban, Mali) had 13-14 caudal peduncle scales. The specimens from Kalaban, Mali will hereafter be referred to as *E.* sp. 'Kalaban'.

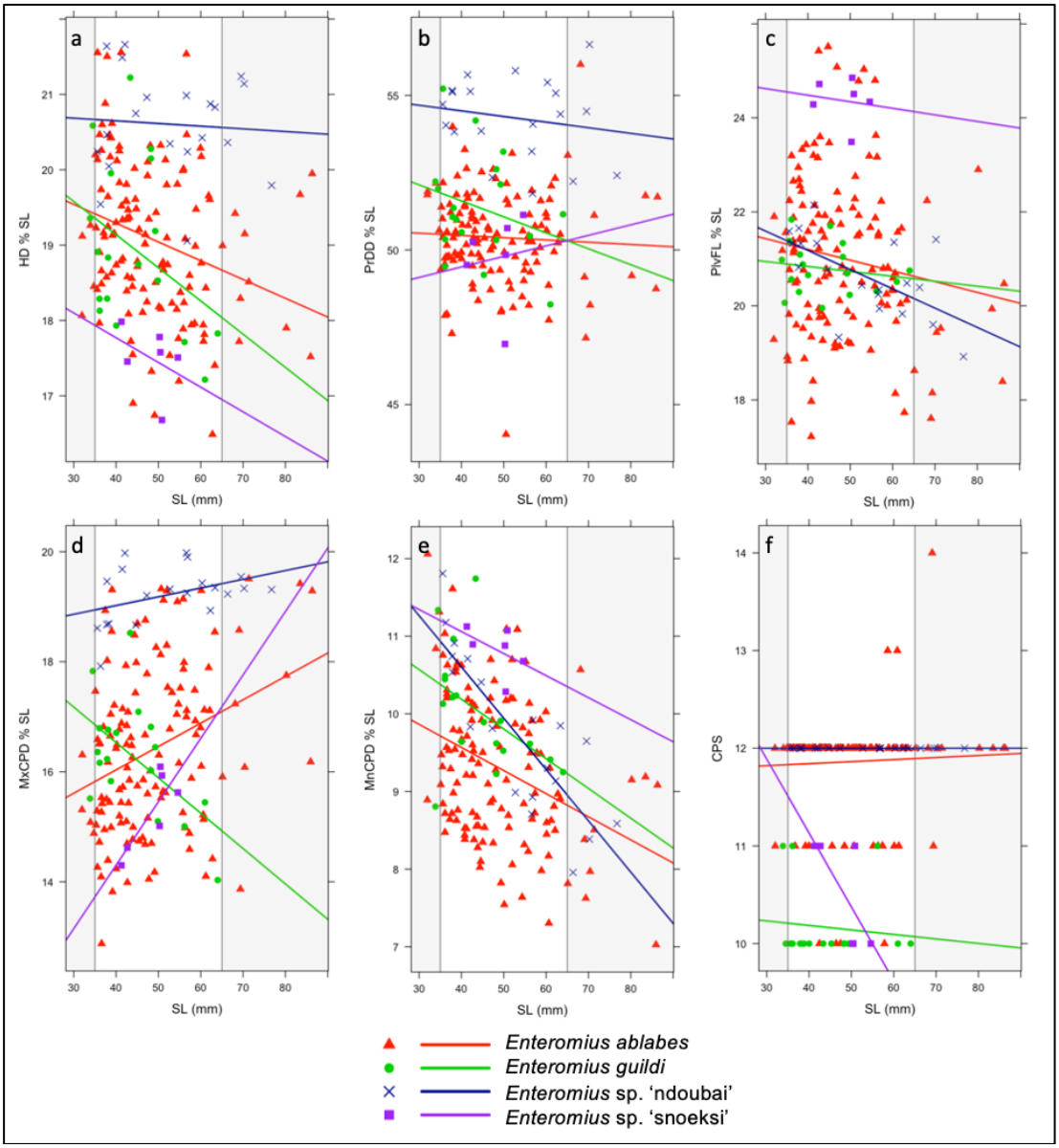

Figure 12: Scatterplot of (a) HD % SL (Head depth) vs. SL, (b) PrDD % SL (Pre-dorsal distance) vs SL, (c) PlvFL % SL (Pelvic fin length) vs SL, (d) MxCPD % SL (Maximum caudal peduncle depth) vs SL, (e) MnCPD % SL (Minimal caudal peduncle depth) vs SL, and (f) CPS (Caudal peduncle scales) vs SL. Examined specimens: 129 specimens of *E. ablabes* (▲), 18 specimens of *E. guildi* (●), 20 specimens of *E.* sp. 'ndoubai' (×), 6 specimens of *E.* sp. 'snoeksi' (■). The indicated window of view is between 35 mm of SL and 65 mm of SL.

Species complex: *Enteromius ablabes* and *Enteromius* sp. 'ndoubai'

A PCA was performed on the log-transformed measurements of 129 specimens of *E. ablabes* and 20 specimens of *E.* sp. 'ndoubai' (Addendum E). Another PCA on the % SL measurements of the same specimens was performed (Figure 13). We chose this PCA to illustrate as specimens from Little Scarcies River (DB157-DB159) had a remarkable position. These specimens, labelled as *E. ablabes* in our analyses, were falling into the morphospace of *E.* sp. 'ndoubai'. These specimens are circled in Figure 13. This observation was less prominent in the results of the PCA on the log-transformed measurement. In both (% SL and log-

transformed) PCAs, the morphospace of *E.* sp. ‘ndoubai’ made a 50% overlap with the morphospace of *E. ablabes*.

A combination of PC1 and PC2 was used where PC1 explains 28.5% of the variance and PC2 21.0% of the variance. No further distinction between the two groups was made on subsequent axes (not illustrated). The highest loadings on PC1, in decreasing order, are PrAD, PoDD2 and PrDD. The highest loadings on PC2, in decreasing order, are DoFL, BD2 and PlvFL.

For the examination of the geographical distribution, the log-transformed results were used, showing PC2 (1.7%) and PC3 (1.1%) (Figure 14). The highest loadings for PC2, in decreasing order, are MnCPD, MxCPD, DoFL and BD2. The highest loadings for PC3, in decreasing order, are PoAD2, ED, PoDD2 and AnFBL. When the groups were divided by their degrees of longitude, a less prominent geographic distribution than in Figure 11 is observed (Figure 14). Similar patterns were found when exploring the measurements expressed by the percentage on SL (not illustrated).

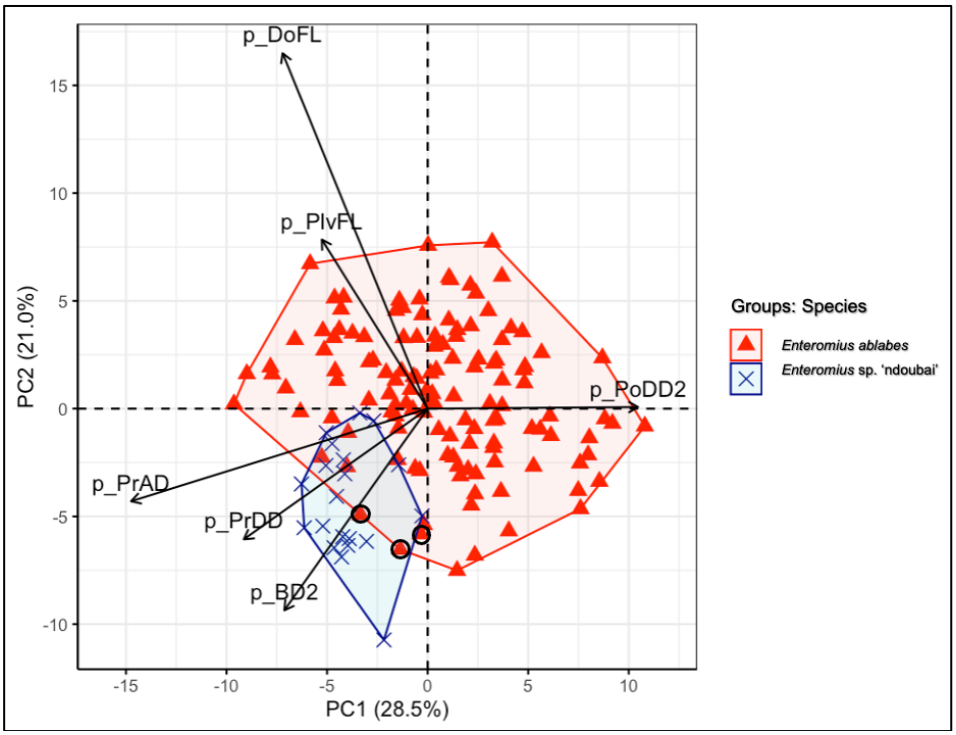

Figure 13: Biplot of a PCA on the % SL measurement of 129 specimens of *E. ablabes* (▲) and 20 specimens of *E.* sp. ‘ndoubai’ (×). Circled data points are specimens from Little Scarcies River (DB157-DB159). The six most important loadings of PC1 and PC2 are indicated with arrows.

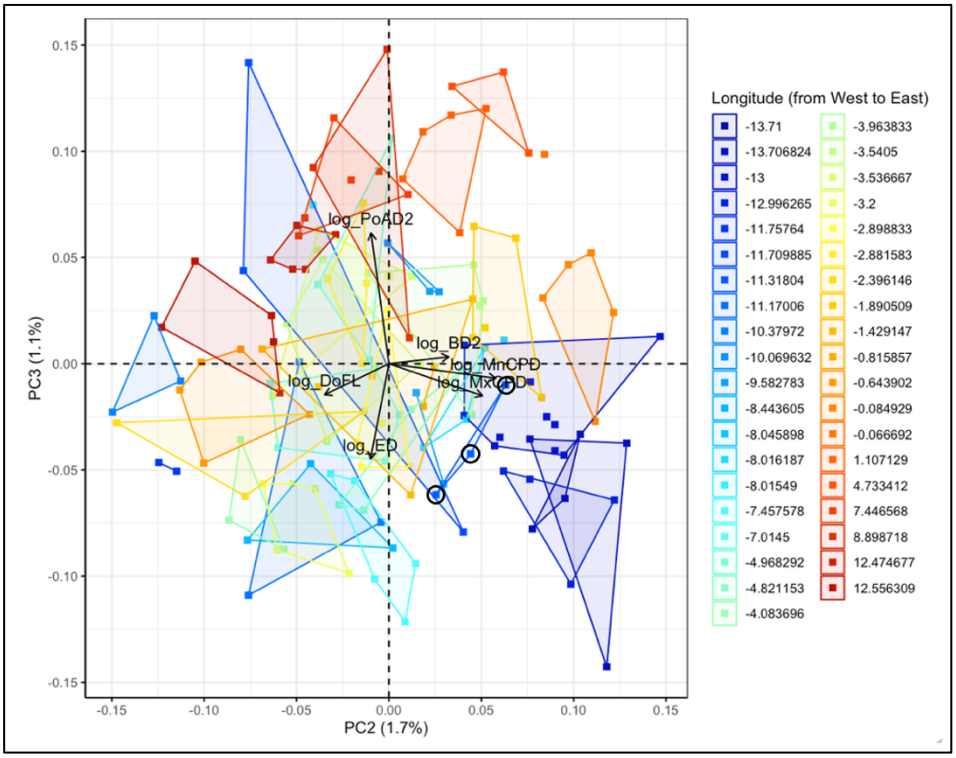

Figure 14: Biplot of a PCA on the log-transformed measurements of *E. ablabes* (n=129) and *E.* sp. 'ndoubai' (n=20) labelled with the corresponding degrees of longitude of the sampling location. Circled data points are specimens from Little Scarcies River (DB157-DB159). The six most important loadings of PC2 and PC3 are indicated with arrows.

Furthermore, the result of the PCA on the meristics of 129 specimens of *E. ablabes* and 20 specimens of *E.* sp. 'ndoubai' using PC1 and PC2 is shown in Figure 15. PC1 explains 14.8% of the variance and the highest loadings, in decreasing order, is the code for the anterior barbel, the code of the posterior barbel and CPS. PC2 explains 11.8% of the variance and the highest loadings, in decreasing order, are LLS_total, GrLow, PDSc and GrUp. Specimens from the two groups could not be distinguished from each other. The other principal components of the PCA were also not able to distinguish the two groups (not illustrated). No clear geographic distribution was observed (Figure 16). Five specimens, situated on the positive part of PC1 and the negative part of PC2, were clearly distinct from the other specimens (circled in green in figure 15 and 16). These were caught in the Niger River in Kalaban (Mali) and showed a scale formula of 4.5/ 28-30 / 3.5; 2.5; 14 (12-13). As observed in the PCA on the % SL measurements, specimens labelled as *E. ablabes* in our analyses sampled from Little Scarcies River were falling into the morphospace of *E.* sp. 'ndoubai'. These specimens are circled in black in figures 15 and 16. These results were similar to the results of the PCA with all examined specimens.

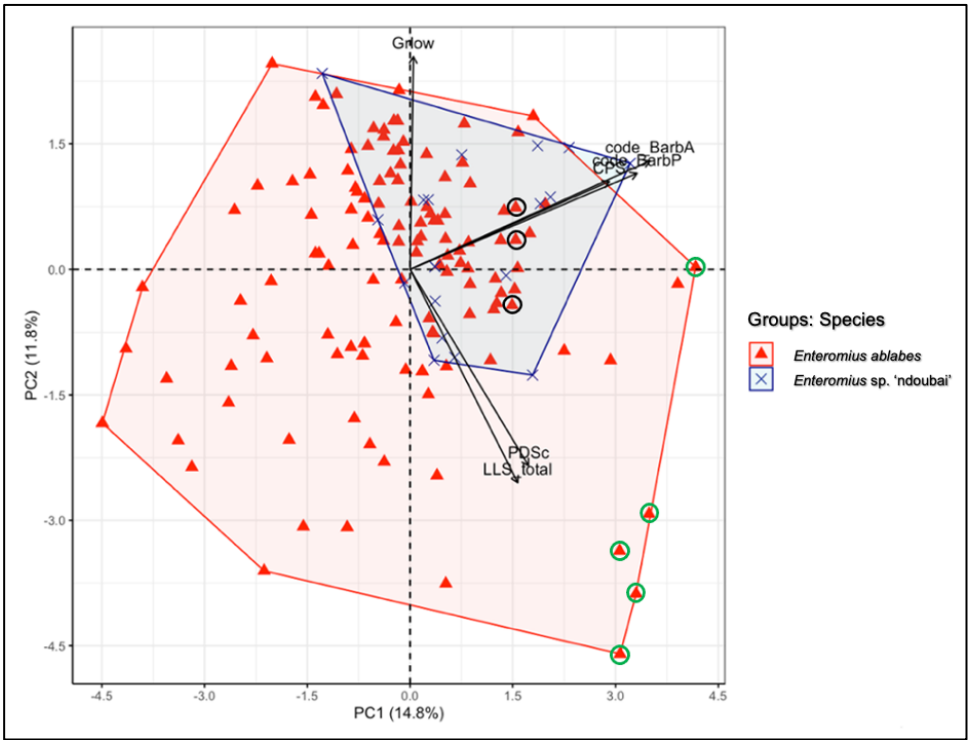

Figure 15: Biplot of a PCA on the meristics of 129 specimens of *E. ablabes* (▲) and 20 specimens of *E.* sp. 'ndoubai' (✕). Circled data points in black are specimens from Little Scarcies River (DB157-DB159), circled in green are *E.* sp. 'Kalaban'. The six most important loadings of PC2 and PC3 are indicated with arrows.

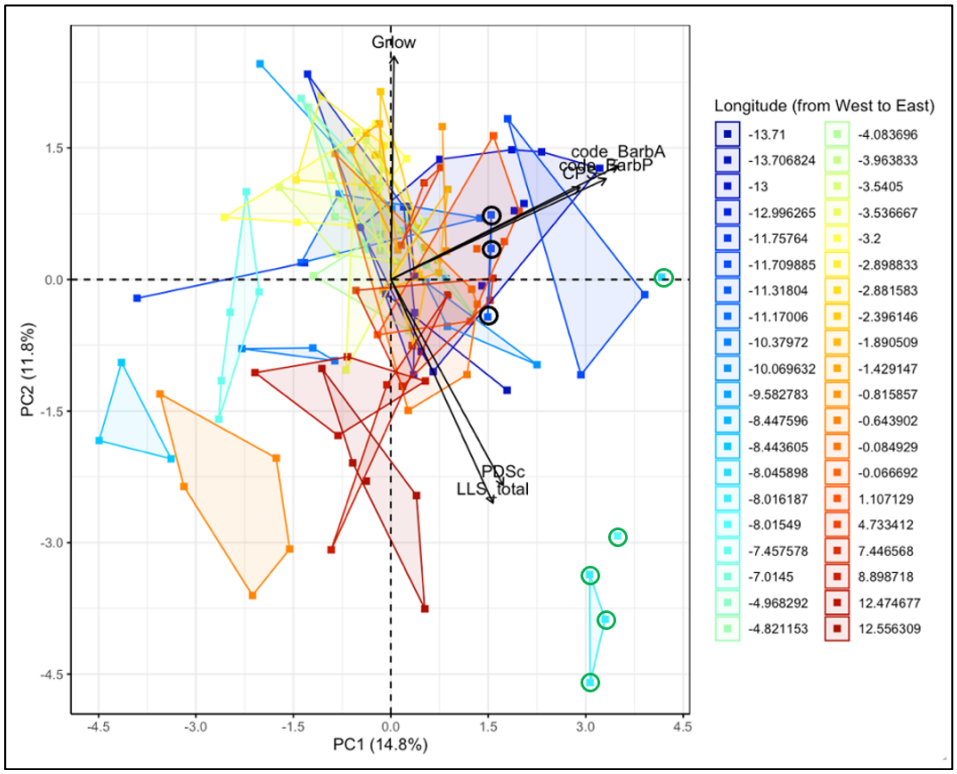

Figure 16: Biplot of a PCA on the meristics of *E. ablabes* (n=129) and *E.* sp. 'ndoubai' (n=20). Groups labelled with the corresponding degrees of longitude of the sampling location. Circled data points in black are specimens from Little Scarcies River (DB157-DB159), circled in green are *E.* sp. 'Kalaban'. The six most important loadings of PC1 and PC2 are indicated with arrows.

Species complex: *Enteromius guildi* and *Enteromius* sp. 'snoeksi'

For the examination of the species complex of *E. guildi* and *E.* sp. 'snoeksi', a PCA was performed on the log-transformed measurements of 18 specimens of *E. guildi* and 6 specimens of *E.* sp. 'snoeksi', showing PC2 (4.6%) and PC3 (1.2%) (Figure 17). The highest loadings for PC2, in decreasing order, are IOW, SnL, PlvFL and AFL. The highest loadings for PC3, in decreasing order, are ED, AFL, SnL, MnCPD and MxCPD. A clear distinction between the two groups was found, based on PC2. PC4 and further principal components were not able to distinguish the two groups (not illustrated). No geographic distribution could be found, based on the degrees of longitude of the sample location of the specimens (Figure 18). The same results were found when exploring the measurements expressed by the percentage on SL (not illustrated).

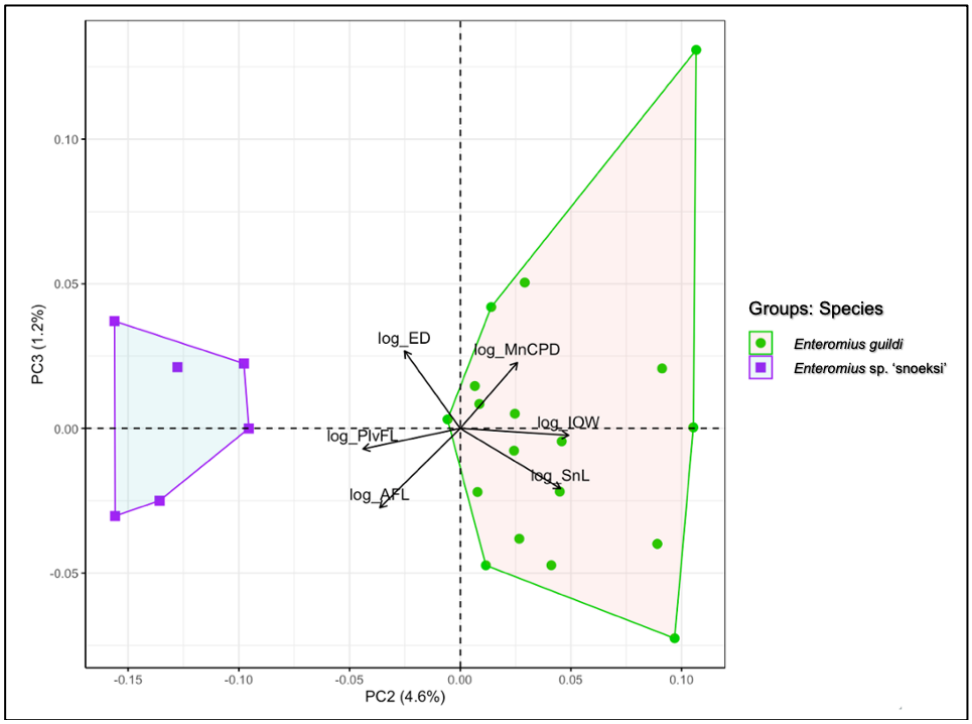

Figure 17: Biplot of a PCA on the log-transformed measurements of 18 specimens of *E. guildi* (●) and 6 specimens of *E.* sp. 'snoeksi' (■). The six most important loadings of PC2 and PC3 are indicated with arrows.

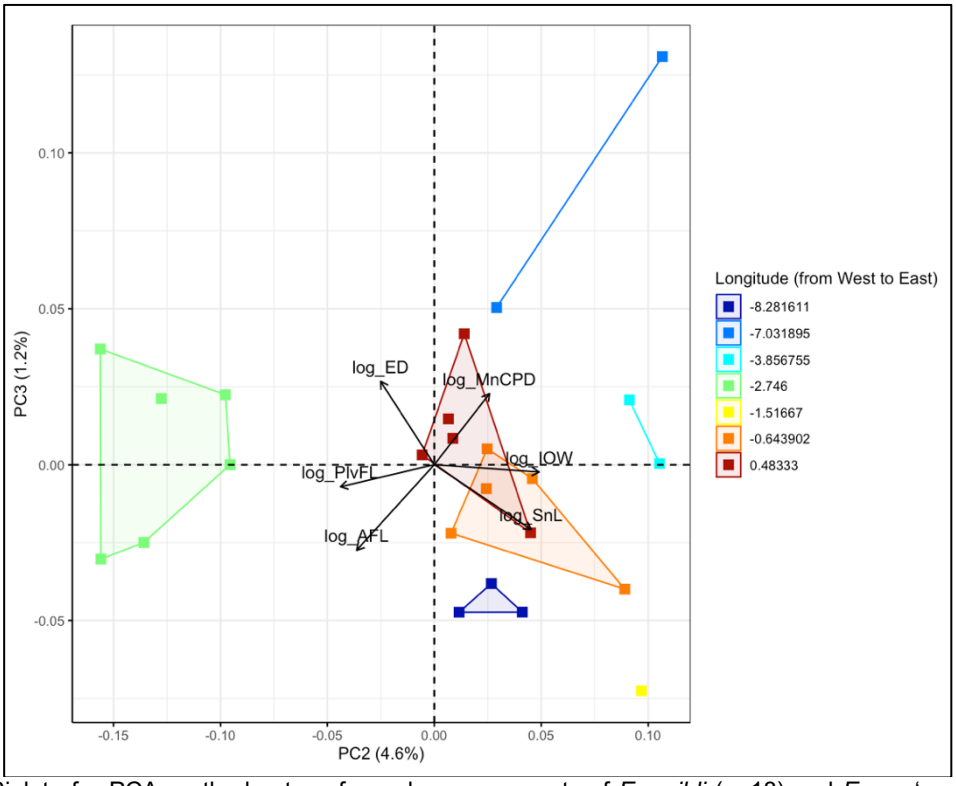

Figure 18: Biplot of a PCA on the log-transformed measurements of *E. guildi* (n=18) and *E.* sp. 'snoeksi' (n=6). Groups labelled with the corresponding degrees of longitude of the sampling location. The six most important loadings of PC2 and PC3 are indicated with arrows.

Another PCA was performed on the meristics of 18 specimens of *E. guildi* and 6 specimens of *E.* sp. 'snoeksi' (Figure 19). PC1 explains 23.6% of the variance and the highest loadings, in decreasing order, are code for the posterior barbel, PDSc, LLS_total and code of the anterior barbel. PC2 explains 16.7% of the variance and the highest loadings are CPS, DFR_branched, PelFR_branched and GrUp. No distinction between the two groups was observed. The other PCs were also not able to distinguish *E.* sp. 'snoeksi' from *E. guildi* (not illustrated). No geographic distribution was found when the groups were divided by the degrees of longitude (Figure 20). Similar results were found when examining the PCA with all studied specimens.

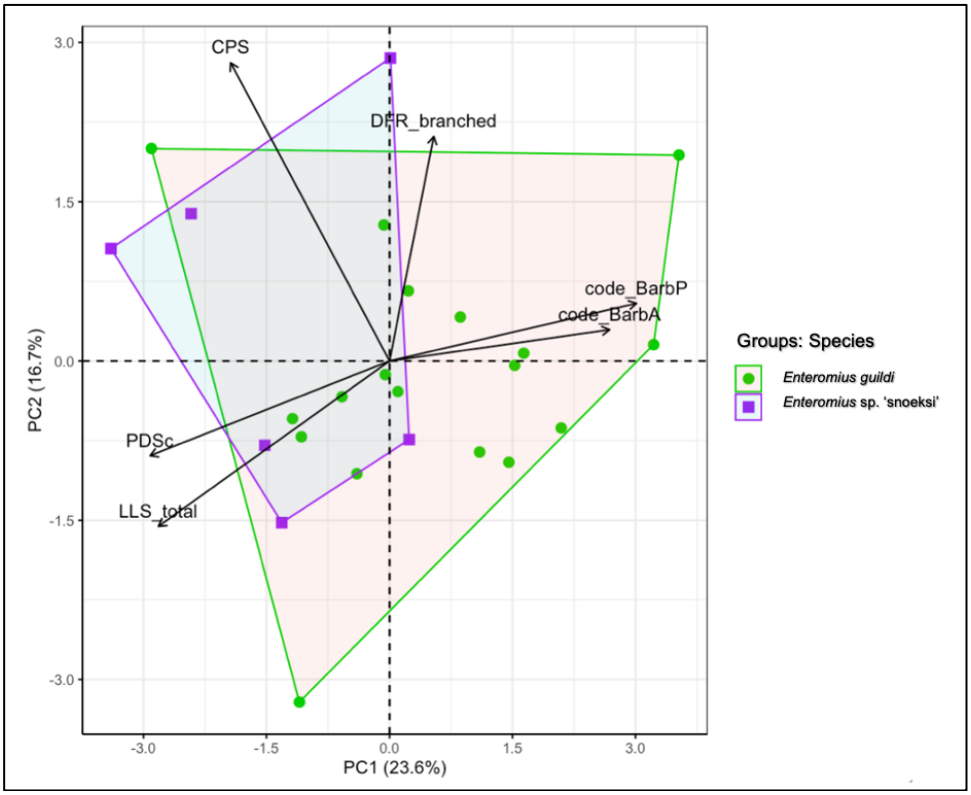

Figure 19: Biplot of a PCA on the meristics of 18 specimens of *E. guildi* (●) and 6 specimens of *E.* sp. 'snoeksi' (■). The six most important loadings of PC1 and PC2 are indicated with arrows.

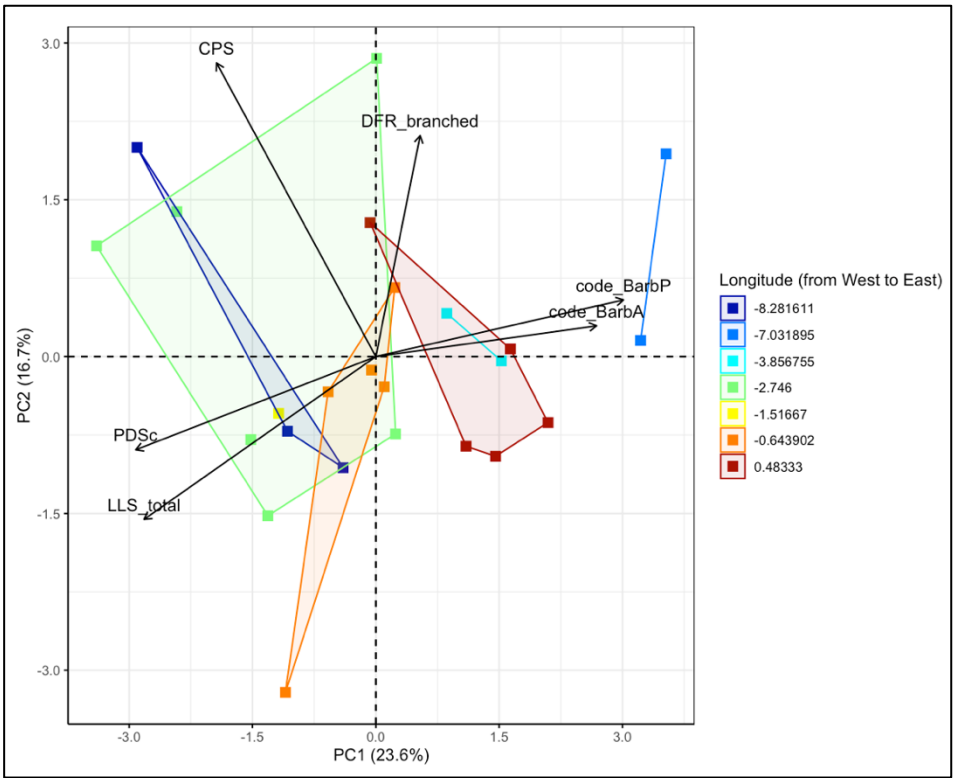

Figure 20: Biplot of a PCA on meristics of *E. guildi* (n=18) and *E.* sp. 'snoeksi' (n=6). Groups labelled with the corresponding degrees of longitude of the sampling location. The six most important loadings of PC1 and PC2 are indicated with arrows.

Comparison between *Enteromius ablabes* and *Enteromius guildi*

Apart from the two species complexes, we compared the two valid species: *E. ablabes* and *E. guildi*. A PCA was performed on the log-transformed measurements of 129 specimens of *E. ablabes* and 18 specimens of *E. guildi* (Figure 21). In the PCA, a distinction between specimens examined by both Bamba (2012) and us (YES) and specimens not examined by Bamba (2012) (NO) was made, using PC2 (1.7%) and PC3 (1.2%). The highest loadings for PC2, in decreasing order, are PoAD2, ED, PoDD2 and SnL. The highest loadings for PC3, in decreasing order, are MxCPD, AnFBL, MnCPD. PC4 and further principal components were neither able to distinguish the different groups (not illustrated). The same results were found when exploring the measurements expressed by the percentage on SL (not illustrated).

However, a last PCA on the meristics showed an interesting insight into our data (Figure 22). A distinction between *E. ablabes* specimens examined by both Bamba (2012) and us (YES) and *E. guildi* is present, while an overlap of the morphospace of *E. guildi* with *E. ablabes* (NO) was observed. For this PCA, we used PC1 (19.3%) and PC2 (14.7%) The highest loadings for PC1, in decreasing order, are code_BarbA, code BarbP, CPS and LLS_total. The highest loadings for PC2, in decreasing order, are LLS_total, PDSc, Grlow, PelFR_branched, PecFR_branched and GrUp.

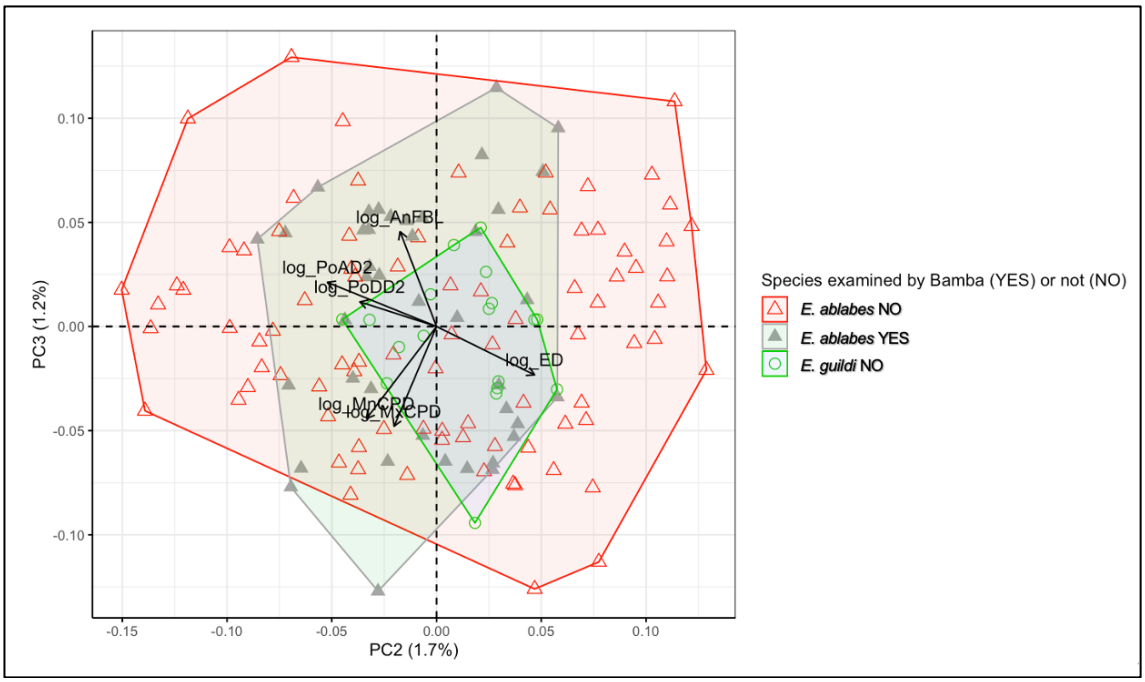

Figure 21: Biplot of a PCA on the log-transformed measurement of 88 specimens of *E. ablabes* not examined by Bamba (2012) (△), 41 specimens of *E. ablabes* examined by both Bamba (2012) and us (▲) and 18 specimens of *E. guildi* not examined by Bamba (2012) (○). The six most important loadings of PC2 and PC3 are indicated with arrows.

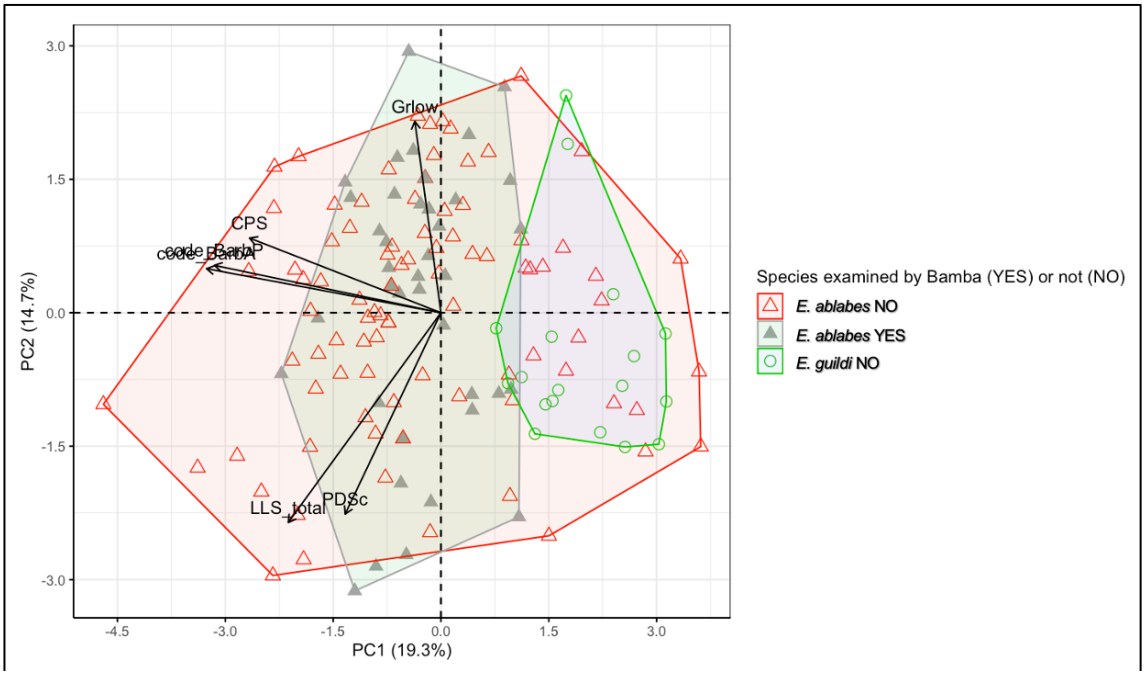

Figure 22: Biplot of a PCA on the meristics of 88 specimens of *E. ablabes* not examined by Bamba (2012) (△NO), 41 specimens of *E. ablabes* examined by both Bamba (2012) and us (▲YES) and 18 specimens of *E. guildi* not examined by Bamba (2012) (○NO). The six most important loadings of PC1 and PC2 are indicated with arrows.

3.2. Genetic results

3.2.1. DNA extractions

Each batch had samples from all four classes of DNA concentrations as described in the material and methods section 2.4.1. (Figure 23). Namely, 28% of the samples (n=17) had a DNA concentration under 0.550 ng/µL, 32% of the samples (n=19) had a DNA concentration between 0.550 ng/µL and 1.000 ng/µL and 40% of the samples (n=24) had a DNA concentration above or equal to 1.000 ng/µL. All negative controls were systematically too low to quantify the DNA concentration with the Qubit 4 Fluorometer. Overall, DNA was not highly concentrated, with the following median values for the different batches: 0.692 ng/µL for batch 1, 0.887 ng/µL for batch 2, 0.758 ng/µL for batch 3, and 0.954 ng/µL for batch 4. Some DNA extractions were particularly concentrated when compared to the overall concentration among different extractions (Figure 23). DNA extractions from samples DB28 and DB49 (not shown) had high concentrations of 8.36 ng/µL and 25 ng/µL respectively. Otherwise, the samples from some specific collector/year resulted in better extracts. The extraction from specimens from the collection of D. Thys van den Audenaerde (1965) performed well. Samples DB28, DB30, DB3, DB41, and DB70 from that collection had DNA concentrations above 2.000 ng/µL, while DB77 and DB79 had concentrations above 1.000 ng/µL. Detailed information about all DNA extracts can be found in Addendum F.

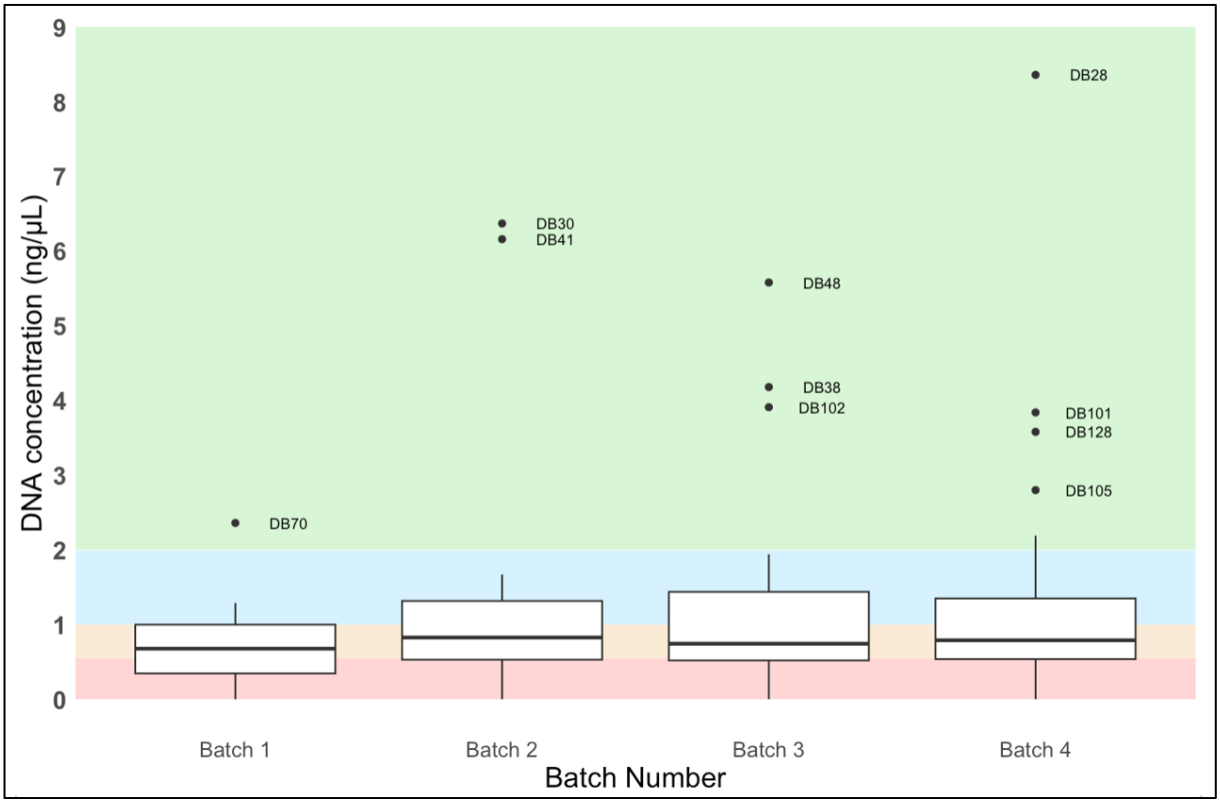

Figure 23: Overview of the DNA extraction results using boxplots. The DNA concentration (y-axis) was plotted against the batch number (x-axis). The background colouring is based on the DNA concentration classes: red for "Very Low", orange for "Low", blue for "Okay"; and green for "High". Samples from the class good are labelled in the plot. In total (=60): 17 samples (28%) are recognized as "Very Low", 19 samples (32%) as "Low", 12 (20%) as "Okay"; and 12 (20%) as "High". Sample DB49 (25 ng/µL) is not shown in the boxplot. The mean of batch 1 (n=6) is 0.954 ng/µL ± 0.773 ng/µL; the mean of batch 2 (n=10) is 1.899 ng/µL ± 2.335 ng/µL; the mean of batch 3 (n=15) is 1.583 ng/µL ± 1.625 ng/µL; and the mean of batch 4 (n=29) is 2.183 ng/µL ± 4.681 ng/µL. Sample IDs can be checked in Addendum F. To increase the readability of the figure, DB49 (Batch 4) is not shown in the figure, as having a high DNA concentration (25 ng/µL). Boxplots were made in RStudio using the 'ggplot2' package.

3.2.2. Sequencing

PCR outputs

The results of the gel electrophoreses are shown in Figure 24. Most of the PCR products for all DNA concentration classes had bands at around 100 bp – 200 bp. Those profiles were compatible with a proper amplification of the different mini-barcodes. The PCR negative control (C-) for the "two-time" PCRs had no bands, while the positive control (C+) for these PCRs gave a clear band at the expected sizes. In contrast, most of the negative controls at the DNA extraction step (c-B1, c-B2, c-B3 and c-B4) exhibited bands on the electrophoreses as did the PCR negative controls (C-). Surprisingly, the samples DB28, DB30, DB38 and DB41 had a high concentration of DNA (4.18 ng/µL – 8.36 ng/µL) (Figure 23) but did not exhibit bands on the electrophoresis (Figure 24). The positive controls (C+) always showed a clear band.

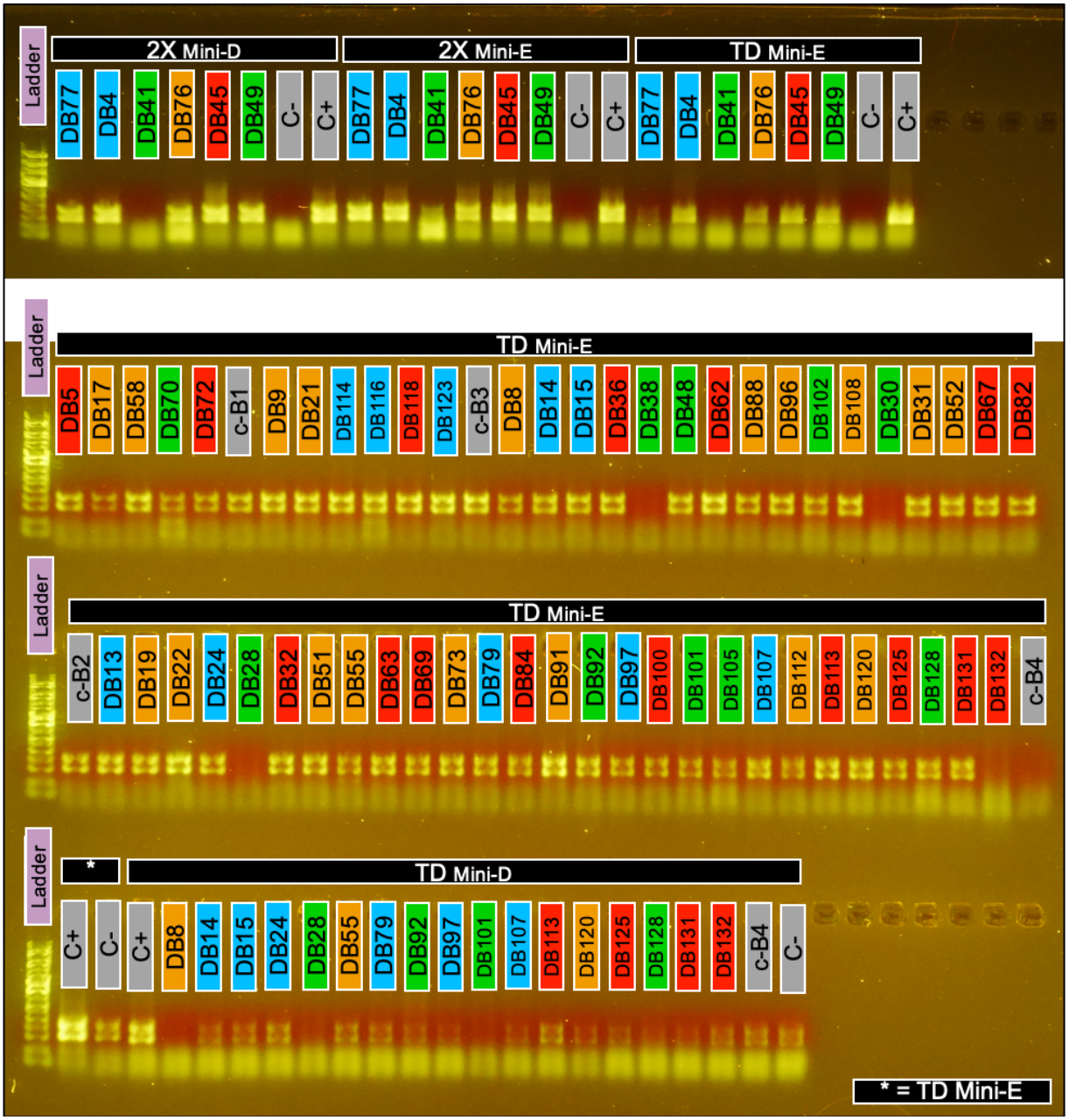

Figure 24: Results of the gel electrophoreses of the PCR samples. “2X”= “two-time” PCR and “TD”= touchdown PCR. FastGene 100 bp DNA Ladder is (Nippon Genetics) used in the gel electrophoreses, indicated with the label “Ladder”. The colouring of the bars is based on the DNA concentration classes: red for “Very Bad”, orange for “Bad”, blue for “Ok”, green for “Good”, and grey for the positive and negative controls. More details on the sample IDs can be found in Addendum C.

Sanger sequencing

The results of the 96 sequences (raw sequences, forward and reverse sequences) blasted in GenBank are given in Table 9. A detailed table with results for every sample can be found in the Addendum G. For most of the PCR products from our DNA extracts, without negative and positive controls (73%) BLAST indicated that sequences matched mostly with *Homo sapiens* sequences. Some sequences (n=6) matched with other organisms such as *Oligotricha lapponica* (Hagen, 1864), *Aseraggodes whitakeri* (Woods, 1966) *Pipa pipa* (Linnaeus, 1758)

and *Sus scrofa* (Linnaeus, 1758) (Addendum G). All sequences from the negative controls of the DNA extractions (c-B#) matched with *Homo sapiens* sequences. The sequences from the PCR negative controls were all labelled as 'BAD' and 60% of them provided no match with Blast. However, one of the PCR negative controls (C-, Touchdown PCR, Mini-E sequence in Addendum G) resulted in a match with an *Enteromius* species, *E. pellegrini* (Poll, 1939). All sequences of the positive controls matched with *Hydrocynus*. The best matches however were with *Hydrocynus vittatus* (Castelnau, 1861) and not with the expected species, *Hydrocynus forskahlii*. For all samples that matched with *Homo sapiens*, the Percent Identity, Max Score and E-value are given (Table 9). A low E-value, a high Query cover and a high Percentage Identity suggested that the match resulted from Blast was of high quality. Most of the samples that matched *Homo sapiens* were of relatively high quality, but the Query cover was low (Table 9).

Table 9: BLAST outcome of the raw sequence data.

Programme	**Number of 'BAD' sequences**	**Number of 'OK' sequences**	**Blast result as *Homo sapiens***	**Query cover ***	**Percent Identity ***	**Max Score ***	**E-value ***
(mini-D) – "two-time" PCR (n=12)	7 (58%)	5 (42%)	5 (42%)	41.4% ± 5.3%	97.2% ± 1.9%	342.4 ± 35.2	1e-73 ± 2e-73
(mini-E) – "two-time" PCR (n=12)	5 (42%)	7 (58%)	9 (75%)	77.6% ± 13.2%	91.9% ± 5.4+	320.1 ± 60.7	6e-67 ± 3e-66
(mini-D) - Touchdown PCR (n=16)	11 (69%)	5 (31%)	11 (69%)	48.2% ± 25.4%	93.5% ± 4.0%	318.4 ± 74.5	2e-50 ± 6e-50
(mini-E) - Touchdown PCR (n=26)	10 (38%)	16 (62%)	23 (88%)	52.0% ± 31.5%	95.3% ± 4.4%	380.3 ± 79.0	9e-46 ± 4e-45
Negative controls extractions (c-B#) (n=10)	3 (30%)	7 (70%)	10 (100%)	41.2% ± 34.8%	96.9 % ± 2.9%	402.8 ± 55.7	2e-75 ± 6e-75
Negative controls PCR (C-) (n=10)	10 (100%)	0 (0%)	1 (10%)	37%	85%	198	8e-46
Positive controls PCR (C+) (n=10)	0 (0%)	10 (100%)	0 (0%)	na	na	na	na

The * indicate the metrics of the blasting outcome for the samples with Human DNA. For those metrics, we provided the means and the standard deviations

The blasting results of the corrected and trimmed sequences that were labelled as “OK” are shown in Table 10. Apart from four sequences, all resulted in *Homo sapiens*. For three sequences there was no blasting result found in GenBank. The TD with Mini-D for sample DB55 resulted in Eukaryotic synthetic construct, but the same metrics were found for *Homo sapiens* (not shown). The length of the sequences was the expected length and the metrics of the blast were all good matches (Table 10).

Table 10: Results of the corrected and trimmed sequences from the blasting in GenBank.

Programme	**Specimen**	**contig?**	**Best match Blast on GenBank (name organism - accession number)**	**Metrics of Blast on GenBank (Query cover - Percent Identity - Max Score - E-value)**	**Length (bp)**
(mini-E) - TD	c-B2	Yes	*Homo sapiens* - OQ731991.1	100% - 100% - 392 - 1e-104	212
(mini-E) - TD	DB77	Yes	*Homo sapiens* - MN176260.1	100% - 100% - 392 - 1e-104	212
(mini-E) - TD	c-B3	Yes	*Homo sapiens* - OQ731991.	99% - 99% - 379 - 1e-100	212
(mini-E) - TD	DB62	Yes	*Homo sapiens* - OQ731991.1	100% - 100% - 392 - 1e-104	212
(mini-E) – 2X	DB76	Yes	Not found	na	216
(mini-E) - TD	DB76	Yes	*Homo sapiens* - OP682652.1	100% - 97% - 366 - 8e-97	212
(mini-E) - TD	DB123	Yes	*Homo sapiens* - MN176260.1	100% - 100% - 392 - 1e-104	212
(mini-E) – 2X	DB45	Yes	*Homo sapiens* - OP682652.1	100% - 99% - 381 - 3e-101	212
(mini-E) - TD	C+	Yes	*Hydrocynus vittatus* - HM882886.1	99% - 99% - 385 - 2e-102	212
(mini-E) - TD	DB91	Yes	*Homo sapiens* - MN176260.1	100% - 100% - 394 - 1e-105	213
(mini-E) - TD	c-B1	Yes	*Homo sapiens* - MN176260.1	100% - 100% - 357 - 4e-94	193
(mini-D) – 2X	C+	Yes	*Hydrocynus vittatus* - LC487217.1	100% - 100% - 363 - 1e-95	196
(mini-E) – 2X	C+	Yes	*Hydrocynus vittatus* - HM882886.1	99% - 99% - 385 - 2e-102	212

(mini-E) - TD	C+	Yes	*Hydrocynus vittatus* - HM882886.1	99% - 99% - 385 - 2e-102	212
(mini-D) - TD	C+	Yes	*Hydrocynus vittatus* - LC487217.1	100% - 100% - 363 - 1e-95	196
(mini-E) - TD	DB58	Forward	Not found	na	207
(mini-E) – 2X	DB4	Forward	Not found	na	165
(mini-D) -2X	DB4	Reverse	*Homo sapiens* - OQ731990.1	100% - 97% - 298 - 2e-76	173
(mini-E) – 2X	DB41	Forward	*Homo sapiens* - MN176260.1	79% - 98% - 316 - 9e-82	228
(mini-E) - TD	DB41	Forward	*Homo sapiens* - MN176260.1	78% - 95% - 289 - 2e-73	222
(mini-E) – 2X	DB41	Reverse	*Homo sapiens* - OP682652.1	94% - 97% - 296 - 9e-76	186
(mini-E) - TD	DB41	Reverse	*Homo sapiens* - OP682652.	98% - 98% - 292 - 1e-74	171
(mini-E) - TD	DB45	Forward	*Homo sapiens* - MN176260.1	100% - 92% - 305 - 2e-78	198
(mini-D) - TD	c-B4	Reverse	*Homo sapiens* - OQ731990.1	100% - 99% - 313 - 8e-81	173
(mini-D) - TD	DB55	Reverse	Eukaryotic synthetic - CP034492.1	100% - 92% - 257 - 4e-64	170
(mini-D) – 2X	DB77	Reverse	*Homo sapiens* - OQ731990.1	100% - 97% - 272 - 1e-68	157
(mini-D) – 2X	DB45	Reverse	*Homo sapiens* - OQ731990.1	98% - 95% - 281 - 2e-71	172
(mini-D) - TD	DB79	Reverse	*Homo sapiens* - OQ731990.1	100% - 98% - 300 - 6e-77	170
(mini-D) – 2X	DB76	Reverse	*Homo sapiens* - OQ731990.1	98% - 98% - 276 - 1e-69	158
(mini-D) – 2X	DB49	Reverse	*Homo sapiens* - OQ731990.1	100% - 97% - 287 - 5e-73	165
(mini-D) - TD	DB113	Reverse	*Homo sapiens* - OQ731990.1	100% - 100% - 320 - 5e-83	173
(mini-E) - TD	DB5	Yes	*Homo sapiens* - MN176260.1	99% - 100% - 418 - 2e-112	227
(mini-E) - TD	DB22	Yes	*Homo sapiens* - MW464147.1	99% - 100% - 388 - 2e-103	212
(mini-D) - TD	c-B4	Reverse	*Homo sapiens* - OQ731990.1	100% - 99% - 283 - 6e-72	157

2X = “two-time” PCR, TD = touchdown PCR. The specimen IDs can be found in Addendum C.

4. Discussion

The discussion of this study is structured similarly to the previous chapter of the thesis. We will go first over the findings on morphometry and second, on the molecular results. After, we will discuss how to bridge the two approaches in order to make the main objective, the exploration of the hidden diversity within *Enteromius* in West Africa, possible. As last, we will discuss the difficulties we encountered using formalin-fixed ethanol-preserved specimens and make a conclusion.

4.1. Morphometry

4.1.1. The *Enteromius ablabes* and *Enteromius* sp. 'ndoubai' species complex

We found significant differences in morphological traits between specimens of *E. ablabes* and *E.* sp. 'ndoubai' based on the MWU tests (Table 8). For the dorsal fin, both the pre-dorsal (PrDD) (Figure 12b) and post-dorsal (PoDD2) distances are significantly different between the two species while differences in the length of the base of the dorsal fin (DFBL) and the length of the dorsal fin itself (DFL) are not significant (Table 8). These findings show that the size of the dorsal fin is similar but the fin has a more posterior position in *E.* sp. 'ndoubai'. The same trend is observed for the anal fin which is positioned more posteriorly in *E*. sp. 'ndoubai' but has proportionately the same size in the two species (Table 8). In addition to the length of the caudal peduncle, resulting from the position of the dorsal and anal fin, also differences in the maximal and minimal caudal peduncle depth (MxCPD and MnCPD) between the specimens of *E. ablabes* and *E.* sp. 'ndoubai' were observed (Table 8 and Figure 12d, 12e). *Enteromius* sp. 'ndoubai' systematically has a deeper caudal peduncle distance. Furthermore, a significantly deeper body depth and head depth (BD2 and HD) is observed in *E*. sp. 'ndoubai' than in *E. ablabes* (Table 8 and Figure 12a). From the findings of the differences in the caudal peduncle, the head and body depth, we conclude that the body of *E. ablabes* is more elongated than the body shape of *E.* sp. 'ndoubai'. Even though the morphospace of both species overlap (Figure 13), it supports the idea that *E.* sp. 'ndoubai' is a valid species. Bamba (2012) found even more significant differences when applying MWU tests (e.g. in eye diameter (ED) and the length of the dorsal fin (DoFL)). When covering a more balanced distribution of *E. ablabes* in our research, the other significant differences mentioned by Bamba (2012) could not be validated.

The meristics also displayed some significant differences within this species complex (Table 8). A lower amount of lateral line scales (LLS_total) and a longer anterior barbel (code_BarbA) in *E.* sp. 'ndoubai' was found (Table 8). Bamba (2012) already observed these differences within the species complex. Significant differences, after Bonferroni correction, for the

branched fin rays in the anal and pectoral fins (AFR_branched and PecFR_branched) are new insights into the species complex (Table 8). When exploring these results, we discovered that some specimens of *E. ablabes* (4.7%) had six branched anal fin rays, but most had seven of these structures. Seven branched anal fin rays are the same amount as most of the specimens of *E.* sp. 'ndoubai' (70%). The mean for the number of branched pectoral fin rays of *E.* sp. 'ndoubai' is significantly higher, however, the majority of specimens (80%) have the same number of these fin rays (14-15) as *E. ablabes*. Thus, we may conclude that the significant differences of the fin rays (AFR_branched and PecFR_branched) are not an accurate indicator of the differences between the two species, while the significant differences of the lateral line scales and the anterior barbel are. The results of the PCA on the meristics support this statement as an overlap was observed (Figure 15).

Geographic distribution of *E.* sp. 'ndoubai'

According to Bamba (2012), specimens of *E.* sp. 'ndoubai' are endemic to the Kakrima River (Konkouré basin). The Konkouré system is located in the Northern Upper Guinea ecoregion (nr 511 in Figure 1) within the Upper Guinea bioregion (Thieme et al., 2006) (Figure 1). According to Thieme et al., (2006), this ecoregion is well recognized for having a high number of endemic species (35-58 endemics) and a high percentage of endemism fish species (23%-31%) in the freshwater rivers. By expanding the research area, we found a group of specimens from Marela, (Little Scarcies River, Republic of Guinea), identified as *E. ablabes* but showing higher similarities with *E.* sp. 'ndoubai' (Figure 13, 15). These specimens were not examined by Bamba (2012). When we re-identified the specimens as *E.* sp. 'ndoubai' and included them in the MWU tests, no other significant differences were found (not illustrated). The labelling of *E. ablabes* for these specimens was therefore maintained in our analyses. The results from the PCAs (Figure 13, 15) show that these specimens are situated in the centre of the morphospace of *E.* sp. 'ndoubai'. Although the Little Scarcies River is not part of the Konkouré system, it is geographically near it (Figure 25). Additionally, like the Konkouré system, the location is situated within the Northern Upper Guinea ecoregion (nr 511 in Figure 1), but close to the border (Thieme et al., 2006). From all this information combined, we conclude that these specimens can be identified as *E.* sp. 'ndoubai.

Figure 25: Map of the Western part of Guinea showing the Konkouré River, the Kakrima River and the Little Scarcies River. The additional specimens of *E.* sp. 'ndoubai' are highlighted with a star. Map made in QGIS using DIVA-GIS shapefiles to display the river systems.

All of the previously mentioned findings result in the conclusion that *E.* sp. 'ndoubai' is a distinct species from *E. ablabes*, therefore being a valid species. Providing a formal description with diagnoses of this species lies outside the scope of this research. *Enteromius* sp. 'ndoubai' is not endemic to the Konkouré system, but also occurs in the geographically close Little Scarcies River in Marela and may potentially exist within the whole Northern Upper Guinea ecoregion.

4.1.2. The *Enteromius guildi* and *Enteromius* sp. 'snoeksi' species complex

The results of the measurements from the PCA and MWU tests show a considerable distinction between the two species (Figure 17 and Table 8). Similar to Bamba (2012), we observed a shorter snout (SnL), a smaller interorbital width (IOW) and a longer dorsal fin length (DoFL) in *E.* sp. 'snoeksi' (Table 8). Furthermore, significant differences were found for the head width (HW) (smaller in *E.* sp. 'snoeksi') and pelvic fin length (PlvFL) (larger in *E.* sp. 'snoeksi') (Table 8, Figure 12c). We conclude that the overall body shape of *E.* sp. 'snoeksi' is different from *E. guildi,* with the former having a more narrow body. For the analyses on meristics, no significant differences based on the MWU-test and a complete overlap in the PCA were found, as did Bamba (2012) (Table 8 and Figure 19). More significant differences were found by Bamba (2012) such as the two pair of shorter barbels (code_BarbA and code_BarbP) and larger eye diameter (ED) in specimens of *E.* sp. 'snoeksi'. Although these differences were observed while examining the specimens, they could not be validated in our study as the sample size was too small. While we only examined 18 specimens of *E. guildi* and 6 specimens of *E.* sp. 'snoeksi', the finding of separating the two species matches with the results of Bamba (2012). Both species are found in sympatry in the Tanoé River, Ivory Coast (Bamba, 2012), and have a geographic distribution within the Nilo-Sudan ecoregion in West Africa (Figure 1) (Thieme et

al., 2006). Because of the small sample size and the scope of the current study, a formal description with diagnoses is not provided.

4.1.3. The two valid species *E. ablabes* and *E. guildi*: are they different?

Both *E. ablabes* and *E. guildi* are valid species and can be distinguished by the number of scales around the caudal peduncle (12 in *E. ablabes* and 10 in *E. guildi*) and a lateral line depressed below the dorsal fin in *E. guildi* (vs. horizontal below the dorsal fin in *E. ablabes*) (Loiselle, 1973). Despite the observed differences, the original description of *E. guildi* by Loiselle (1973) already reported close affinities to *E. ablabes*. Results from Bamba (2012) support this assertion. He observed in some populations of *E. guildi* that the caudal peduncle scales (CPS) varied from eight to twelve and that the lateral line is horizontal below the dorsal fin. These findings raised the question of whether the two species are synonyms.

Our highly significant value for the number of caudal peduncle scales in Table 8 confirms the clear difference between *E. ablabe*s and *E. guildi*. Nevertheless, the additional results (Table 8 and Figure 8, 10) raised the question: “Is there a possible synonymy of both species?” In the results of the PCAs on measurements and meristics, the distinction between both species cannot be seen as both morphospaces almost completely overlap each other (Figure 8, 10). Furthermore, from the MWU tests, only the length of the barbels were significantly different (Table 8). Our results confirmed the observations of Bamba (2012), namely the variation in CPS in *E. guildi* (10 or 11) (Figure 12f). Additionally, we found variation in CPS in *E. ablabes* (10 to 14) (Figure 12f). When reviewing the literature, the number of scales around the caudal peduncle appears to be invariable within the species. Bamba (2012) only observed a few specimens of *E. ablabes* (six out of 207) with 10 caudal peduncle scales. He never mentioned observing eleven, thirteen or more caudal peduncle scales in *E. ablabes*. Also in older literature like Lévêque et al., (1990) and Teugels & Hopkins, (2007) specimens of *E. ablabes* and *E. guildi* are reported to have an invariable number of scales around the caudal peduncle (10 in *E. guildi* and 12 in *E. ablabes*). Since re-identification prior to examination of the specimens was not implemented in this study, we conclude that specimens with eleven CPS can be either *E. ablabes* or *E. guildi*. This statement only holds when the specimen is reported within the geographic distribution of *E. guildi*. If the specimen originated from outside the geographic distribution of *E. guildi*, the specimen can be seen as *E. ablabes*.

As we did not re-identify the specimens before the analyses, there is a chance that misidentified specimens were included. In the PCA of the meristics where we divided the groups based on the identification level, we can distinguish the specimens of *E. ablabes* that Bamba (2012) examined (YES) and the *E. guildi* specimens (Figure 22). From this result, we can understand why Bamba (2012) rejected the suggestion of a possible synonymy. However,

the large overlap in measurable characters (Figure 21) and our findings covering the geographic distribution (Figure 9) make room for a debate about a possible synonymy. Still, we conclude that both species are still valid, but require a genetic research for more insights.

4.1.4. *Enteromius* sp. 'Kalaban', the specimens from Kalaban, Mali

Even though *E. ablabes* and *E.* sp. 'ndoubai' cannot be distinguished in the PCA of the meristics (Figure 15), specimens from Kalaban (Mali) are situated mainly in the positive part of the PC1 and the negative part of PC2 (Figure 15, 16). When examining the specimens, more scales on the lateral line (26-30), around the caudal peduncle (12-14) and between the dorsal fin and lateral line (4.5) were observed. As specimens are different from the other specimens examined in this study, we consider them as a distinct species, hereafter referred to as *E.* sp. 'Kalaban'.

Kalaban is a part of Bamako, the capital of Mali, located in the southwestern part along the Niger River. It is situated in the Upper Niger ecoregion and Nilo-Sudan bioregion (Thieme et al., 2006). The specimens of *E.* sp. 'Kalaban' were compared to the species of *Enteromius* reported from the Kalaban region, based on GBIF, using the descriptions in FishBase. Two species, *Enteromius perince* (Rüppel, 1835) and *Enteromius nigeriensis* (Boulenger, 1903) matched closely with *E.* sp. 'Kalaban'. The specimens of *E.* sp. 'Kalaban' could be distinguished from *E. nigeriensis* by their observed SL of almost 70 mm (vs. 50 mm) and a larger number of caudal peduncle scales (12-14 vs. 12) (Froese & Pauly, 2022). *Enteromius perince* was also rejected as this species is characterized by three distinct dots. When using various identification keys (Lévêque et al., 1990; Bamba, 2012) never led us to *E. ablabes* because of the differences in scales. Further research is needed to clarify if these specimens are from a species that remains yet to be described or are members of an already know taxa.

Figure 26: Specimen DB146 from the lot MNHN 2000-0772. The sampling location is in the Niger River in Kalaban, Bamako (Mali). A black band or series of dots was observed on the lateral line scale of the specimen. More information on this specimen can be found in Addendum C.

4.2. Genetic approach

4.2.1. DNA extraction

Using the adapted protocol of Diedericks (2017) we obtained extractions with DNA concentrations falling into the range of other studies (e.g. Ruane & Austin, 2017) for which it was possible to conduct genetic analysis (Figure 23). In addition, we did not detect DNA in our negative controls using a Qubit high-sensitivity assay. Hence, the protocol can be considered suitable to extract DNA out of formalin-fixed specimens. Ruane & Austin, (2017) indicated some success in genetic analysis based on extracted DNA from formalin-fixed specimens with a DNA concentration above 1 ng/µL. Although this work was conducted on snakes, their findings on how to tackle formalin-fixed ethanol-preserved specimens are interesting for ichthyological research. To the best of our knowledge, the relation between the ability to obtain DNA sequences and the DNA concentrations of the extractions from formalin-fixed specimens is rarely reported in the literature, making it difficult to be certain of a positive outcome.

Through the process of obtaining DNA, we observed some potential underlying conditions related to the outcomes of DNA extraction. For example, we did not observe a clear association between the specimens' preservation time and the amount of extracted DNA. This means that an old specimen might perform equally well as a recent one. In our study, specimens collected in the 1960s did perform as well as specimens from early 2000 (Addendum F). We also observed that DNA extraction from samples from the same collection (i.e. specimens collected by the same collector(s) in the same year) are falling roughly in the same classes of concentrations (cf. section 2.4.1 in materials and methods) (Addendum F). For example, the DNA extractions done on samples from the 1965 collection of D. Thys van den Audenaerde always resulted in high concentrations (> 2 ng/µL), while the DNA concentrations for samples from the 2007 collection of Bamba always were below 1 ng/µL (Addendum F). This "collector effect" indicated that the preservation condition associated with each of those collections (e.g. time in formalin, the sort of formalin, etc.) is probably crucial (Chakraborty et al., 2006; Paireder et al., 2013; Hykin et al., 2015). Most of the time, these specifications are not available for our museum specimens. Therefore, it is still challenging to estimate if a given specimen will perform well or not in DNA extraction. It is important to mention that these observations are not statistically tested and remain empirical observations. Furthermore, some aspects of the DNA extracts (e.g. DNA fragmentation and quality) are not investigated in this study.

4.2.2. Universal mini-barcodes, Sanger sequencing and future challenges

Other studies on formalin-fixed fish (Shokralla et al., 2015; Appleyard et al., 2021) and on snake (Ruane & Austin, 2017) showed the feasibility of doing genetic analysis. In that regard, the use of primers to target the COI gene appears to be promising, though challenging

(Shokralla et al., 2015; Appleyard et al., 2021). Still, in our case study, universal primers failed to provide decent sequences, as the methods seem highly sensitive to contamination. When blasting all sequences from samples and negative controls of the extractions, most matched with *Homo sapiens* (Table 9, 10). This result shows that we amplified and sequenced mostly DNA from contaminants instead of fish DNA. As we mostly detect human DNA in the negative controls of the extraction and not in the PCR negative controls, we suspect some contamination to occur during the extraction step. We were not able to quantify DNA in our negative controls of the extraction, showing that if contamination occurs, it is still probably extremely lowly concentrated. Our main hypothesis is that we observe competition in the PCRs between two different kinds of DNA templates. We suspect that this "conflict" between the lowly-concentrated human DNA and highly-degraded fish DNA in the PCR results in the outcompeting of the fish DNA by the human DNA or by differential co-amplification/ PCR chimaeras. Such PCR chimaeras resulted probably in noisy electropherograms (Figure 7E) and unrelated sequences matches with e.g. wild boar (*Sus scrofa*) and the common Surinam toad (*Pipa pipa*) (Addendum G). Additionally, one of the PCR negative controls (C- Mini-E Touchdown PCR) matched with *Enteromius pellegrini.* This contamination is probably due to remnant DNA from other research conducted in the laboratory of RMCA, as ongoing works are conducted on that specific species. As Diedericks (2017) reported, DNA extractions from formalin-fixed tissues should ideally be performed in a laboratory that has not yet been exposed to fish DNA.

From these findings, we conclude that our method of DNA extraction and Sanger sequencing is extremely sensitive to different types of contamination. Our results are consistent with previous work (Appleyard et al., 2021), showing that Sanger sequencing of mini-barcodes is challenging in formalin-fixed specimens (Table 9, 10). In their study, specimens that were in formalin for over eight weeks (referred to as "Ugly") produce messy and unreadable sequences on the electropherograms because of contamination (e.g. exogenous DNA). They also showed that almost 50% of their samples referred to as "Bad" (samples with formalin fixation times between two and seven weeks) were contaminated with exogenous DNA. Using their classification, our samples could probably be categorised into the "Ugly" or "Bad" class.

Two other interesting observations could be drawn from our results. First, we obtain sequences for all the positive controls and they match with the genus *Hydrocynus* (Table 9) for all tested conditions. It indicates that the PCR programmes are correctly set up. Sequences match with another species of *Hydrocynus* (*Hydrocynus vittatus*) but this can be explained by the taxonomic coverage bias in GenBank for this species (Boom, pers. commun.). Second, samples extracted from the specimens of the 1965 collection of D. Thys van den Audenaerde did not amplify while the DNA extractions were promising (> 2 ng/µL) (Addendum F). A possible

explanation is the presence of PCR inhibitors in these samples. Different chemical components linked to the preservation can interfere with the PCRs. Since we are unaware of the preservation steps performed by the collector, it is unknown which chemical(s) functions as PCR inhibitor(s) in this case.

Due to the new knowledge gained during the present study, we foresee many opportunities for future research. Although the extraction method for formalin-fixed ethanol-preserved specimens tested in this work appeared to provide DNA, alternative methods might provide better results and could be worth testing further. In that regard, the phenol-chloroform extraction protocol with treatment with heated alkali buffer solution used in Hykin et al., (2015) seems promising. However, such a method requires certain facilities as it involves the use of chemicals that could only be handled in laboratories with a proper fume hood (e.g. phenol and chloroform). Additionally, our extractions, even when done carefully are exposed to some level of contamination. Working in a dedicated room might also prove beneficial. Some challenges remain, however. Hykin et al., (2015) suggest sampling a large amount (around one gram) of starting material from multiple different tissues to ensure optimal results. This is an extremely destructive way of handling the samples, which cannot be done with type or highly valued specimens. Mini-barcodes obtained using universal primers and Sanger sequencing do not appear as a reliable method to conduct genetic analysis on the DNA extractions obtained in our study. Those results, as mentioned earlier, echo other studies, like Appleyard et al., (2021), showing mixed results using Sanger sequencing. Englmaier et al., (2020b) and Hykin et al., (2015) succeeded by using Sanger sequencing but they used specific primer pairs to amplify targeted sequences. This is probably a robust and valid option but its implementation also leads to other challenges. For example, the design of the specific primers targeting short sequences for species of which one has no prior knowledge is difficult. Other sequencing approaches relying on Illumina sequencing, already resulted in sequences from formalin-fixed specimens (Ruane & Austin, 2017; Appleyard et al., 2021; Hahn et al., 2022). Appleyard et al., (2021) realised Illumina amplicon libraries using mini-barcode primers of Shokralla et al., (2015) and were able to get DNA sequences that otherwise failed using mini-barcode and Sanger sequencing. Using targeted enrichment Illumina libraries (TE) looks also promising. With TE, a large subset of genes specific to the targeted taxonomic group is isolated from genomic DNA before sequencing. Thereby, the specificity of the sequencing is higher, which could lead to the sequencing of numerous different genes from the extracted fish DNA instead of contaminating human DNA (Ruane & Austin, 2017). Collecting samples suitable for genetic purposes (i.e. gin tissue sampled from fresh specimens and stored in 100% ethanol) might be an alternative to formalin-fixed tissue to study the *E. ablabes* and *E. guildi* case. But it would require important financial and logistical resources.

4.3. Making sense of the diversity in *Enteromius*: when genetics and morphology are integrated.

Although we gained new insights into the species complexes and confirmed *E.* sp. 'ndoubai' and *E.* sp. 'snoeksi' as good candidates for valid species, the nature of morphological diversity in the four species and their evolutionary relationships remains elusive. By using genetic evidence, we could test more robustly if our morphometric findings are consistent with *E.* sp. 'ndoubai' and *E.* sp. 'snoeksi' being two new species. In biological systems where morphologies exhibit complex patterns, morphometric approaches alone are generally insufficient to delineate species (Kanu et al., 2022). Thereby, a proper evaluation of the taxonomic status of *E.* sp. 'ndoubai' and *E.* sp. 'snoeksi' might largely benefit from future genetic analysis.

The combination of morphometric and genetic approaches is already known for suggesting pseudo-cryptic diversity in *Enteromius* species (Van Ginneken et al., 2017; Decru et al., 2022; Maetens et al., in prep.) It can provide insights into the undescribed diversity and helps in pinpointing putative new species. Regarding our case study, Kanu et al., (2022) recently identified, based on the cyt b sequences (mtDNA), a new mitochondrial lineage of *E. ablabes* in the Rokel River basin (Liberia, with a genetic distance of 10.1% with other known *E. ablabes* specimens). They concluded that it is probably a new undescribed species. These results echo the findings of Van Ginneken et al., (2017) in the Congo basin. They discovered 23 new lineages of four a priori-defined species of *Enteromius* using a genetic approach targeting the COI gene. From this result, they were able to separate the specimens with different lineages based on a morphometric analysis. They also reported that multiple lineages of the same a priori-defined species co-exist in some rivers. These results suggest considerable underestimated diversity within the *Enteromius* genus across Africa (Van Ginneken et al., 2017), with possible many species yet to be formally identified and described. For East Africa, Decru et al., 2022 found hidden diversity in various cyprinid genera. They already stated that the hidden diversity is less prominent than in the Congo basin as the Congo basin harbours a wide diversity of habitats (Snoeks et al., 2011). The clear geographic distribution found in the PCA combining both studied species complexes (Figure 9) suggests that we found hidden diversity of *Enteromius* in West Africa. As the genetic approach did not provide DNA sequences of interest, the amount of hidden diversity cannot be quantified.

4.4. Problems encountered by working with formalin-fixed specimens

While examining the 173 formalin-fixed ethanol-preserved specimens, we encountered specimens in various conditions of preservation. Some had a rigid structure, while others were very flexible. Furthermore, sometimes barbels, gill rakers or fins were broken or damaged so that specimens could not be measured. The way of preservation and fixation (e.g. use of buffered or unbuffered formalin) and the handling of the specimens on the field is determined by the collector of the specimens. In that regard, our data suggest that the sample condition might influence the amount of extracted DNA but also affects the ability to use the DNA in PCR (Chakraborty et al., 2006; Paireder et al., 2013; Hykin et al., 2015).

Identification and determination of specimens in museum collections can date back some time. Over time, the insights into the species change and with that also the species descriptions. As a lesson learned, in future research, a re-identification should best be made. This is challenging as the colours of the fish are no longer visible or have faded due to the preservation in ethanol, or the overall poor condition of the specimens. In this thesis, no reidentification was done, which may lead to include misidentified specimens in the analyses.

5. Conclusion

The main goal of our work was to explore the hidden diversity of *Enteromius* in West Africa using two species complexes: *E. ablabes*/*E.* sp. ‘ndoubai’ and *E. guildi*/*E.* sp. ‘snoeksi’. Therefore, a morphometric approach was performed in a wide and balanced geographical context. As genetic approaches are highly valuable in detangling the diversity in *Enteromius*, we also tested the feasibility to apply genetic approaches to those complexes using specimens from the RMCA collection.

Through this work, we confirmed that both *E.* sp. ‘ndoubai’ and *E.* sp. ‘snoeksi’ are valid new species. Furthermore, our morphometric results suggest a larger distribution for *E.* sp. ‘ndoubai’ species which is therefore not confined to the Konkouré system. Our results also suggest a putative new species, *E.* sp. ‘Kalaban’. Additional research and revision on specimens of this part of the Niger River have to be conducted to validate this finding.

Universal mini-barcoding combined with Sanger sequencing does not appear robust, as being very sensitive to contamination. Nevertheless, DNA extractions on formalin-fixed ethanol-preserved specimens using the modified protocol of Diedericks (2017) appears to work in the RMCA lab. This DNA extraction method should be further validated using other sequencing methods, either mini-barcode amplicon or targeted-enrichment libraries coupled with Illumina sequencing. It is also may be worthwhile to test alternative DNA extraction protocols.

Future research using genetic evidence either on formalin-fixed specimens or on newly collected material from the field is critical to quantify the pseudo-cryptic diversity suspected to occur in these *Enteromius* species. It is important to provide new and correct insides in the number of species at hand, which is urgently needed to address conservation issues.

6. References

Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., Coad, B., Mandrak, N., Balderas, S. C., Bussing, W., Stiassny, M. L. J., Skelton, P., Allen, G. R., Unmack, P., Naseka, A., Ng, R., Sindorf, N., … Petry, P. (2008). Freshwater ecoregions of the world: A new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, *58*(5), 403–414.

Adams, M., Raadik, T. A., Burridge, C. P., & Georges, A. (2014). Global biodiversity assessment and hyper-cryptic species complexes: More than one species of elephant in the room? *Systematic Biology*, *63*(4), 518–533.

Agnèse, J.-F., Berrebi, P., Lévêque, C., & Guégan, J.-F. (1990). Two lineages, diploid and tetraploid, demonstrated in African species Barbus (Osteichthyes, Cyprinidae). *Aquatic Living Resources*, *3*(4), 305–311.

Appleyard, S. A., Maher, S., Pogonoski, J. J., Bent, S. J., Chua, X. Y., & McGrath, A. (2021). Assessing DNA for fish identifications from reference collections: the good, bad and ugly shed light on formalin fixation and sequencing approaches. *Journal of Fish Biology*, *98*(5), 1421–1432.

Armbruster, J. W. (2012). Standardized measurements, landmarks, and meristic counts for cypriniform fishes. *Zootaxa*, (3586), 8–16.

Bamba, M. (2012). *THE 'BARBUS' SPECIES (CYPRINIFORMES, CYPRINIDAE) FROM CÔTE D'IVOIRE: A SYSTEMATIC REVISION OF SOME WEST AFRICAN SPECIES COMPLEXES*. *Not published*. KULeuven.

Bamba, M., Vreven, E. J., & Snoeks, J. (2011). Description of Barbus teugelsi sp. nov. (Cypriniformes: Cyprinidae) from the little scarcies basin in Guinea, Africa. *Zootaxa*, *65*(2998), 48–65.

Barbarossa, V., Bosmans, J., Wanders, N., King, H., Bierkens, M. F. P., Huijbregts, M. A. J., & Schipper, A. M. (2021). Threats of global warming to the world's freshwater fishes. *Nature Communications*, *12*(1), 1–10.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., & Ferrer, E. A. (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, *471*(7336), 51–57.

Berrebi, P., Kottelat, M., Skelton, P., & Ráb, P. (1996). Systematics of Barbus: State of the art and heuristic comments. *Folia Zoologica*, *45*, 5–12.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, *22*(3), 148–155.

Bleeker, P. (1863). Natuurkundige van de Hollandsche Maatschappij der Wetenschappen te Haarlem, *v. 18114*((Ser. 2)), Pl. 23 (fig. 1).

Bookstein, F., Chernoff, B., Elder, R., Humphries, J., Smith, G., & Strauss, R. (1985). Morphometrics in Evolutionary Biology. *Journal of the Academy of Netural Sciences of Philadelphia*, *Special pu*(15), 1–277.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances*, *1*(5), 9–13.

Chakraborty, A., Sakai, M., & Iwatsuki, Y. (2006). Museum fish specimens and molecular taxonomy: A comparative study on DNA extraction protocols and preservation techniques. *Journal of Applied Ichthyology*, *22*(2), 160–166.

Cowie, R. H., Bouchet, P., & Fontaine, B. (2022). The Sixth Mass Extinction: fact, fiction or speculation? *Biological Reviews*, *97*(2), 640–663.

Cuvier, G., & Cloquet, H. (1816). *Le Règne Animal distribué d'après son organisation pour servir de base à l'histoire naturelle des animaux et d'introduction à l'anatomie comparée. Les reptiles, les poissons, les mollusques et les annélides.* (1st ed.). Paris.

Daget, J., & Iltis, A. (1965). *Poissons de Côte d'Ivoire (eaux douces et saumatres)*. *Mémoires de l'I.F.A.N*. Paris.

Darwall, W., Bremerich, V., De Wever, A., Dell, A. I., Freyhof, J., Gessner, M. O., Grossart, H. P., Harrison, I., Irvine, K., Jähnig, S. C., Jeschke, J. M., Lee, J. J., Lu, C., Lewandowska, A. M., Monaghan, M. T., Nejstgaard, J. C., Patricio, H., … Weyl, O. (2018). The Alliance for Freshwater Life: A global call to unite efforts for freshwater biodiversity science and conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *28*(4), 1015–1022.

Decru, E. (2015). *The icthyofauna of the central Congo basin: diversity, distribution in the north-eastern tributaries*.

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*, 342–352.

Decru, E., Vranken, N., Maetens, H., Mertens De Vry, A., Kayenbergh, A., Snoeks, J., & Van Steenberge, M. (2022). DNA barcoding the Lake Edward basin: high taxonomic coverage of a tropical freshwater ichthyofauna. *Hydrobiologia*, *849*(8), 1743–1762.

Diedericks, G. (2017). *Determining the mechanisms driving the invasion success of the smallmouth bass (Micropterus dolomieu , Lacepède 1802)*.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A. H., Soto, D., Stiassny, M. L. J., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews of the Cambridge Philosophical Society*, *81*(2), 163–182.

Dujardin, J. P. (2017). *Modern Morphometrics of Medically Important Arthropods*. *Genetics and Evolution of Infectious Diseases: Second Edition*. Elsevier Inc.

Englmaier, G. K., Tesfaye, G., & Bogutskaya, N. G. (2020a). A new species of Enteromius (Actinopterygii, cyprinidae, smiliogastrinae) from the awash river, Ethiopia, and the re-establishment of e. akakianus. *ZooKeys*, *2020*(902), 107–150.

Englmaier, G. K., Viñuela Rodríguez, N., Waidbacher, H., Palandačić, A., Tesfaye, G., Gessl, W., & Meulenbroek, P. (2020b). New data on garra makiensis (Cyprinidae, labeoinae) from the awash river (ethiopia) with remarks on its relationships to congeners on the arabian peninsula. *ZooKeys*, *2020*(984), 133–163.

Fricke, R., Eschmeyer, W. N., & Van der Laan, R. (2022). Eschmeyer's catalog offishes: genera, species, references. *Electronic version*. From: http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp [Retrieved

Froese, R., & Pauly, D. (2022). FishBase. *World Wide Web electronic publication*. From: www.fishbase.org [Retrieved

Garrigos, Y. E., Hugueny, B., Koerner, K., Ibañez, C., Bonillo, C., Pruvost, P., Causse, R., Cruaud, C., & Gaubert, P. (2013). Non-invasive ancient DNA protocol for fluid-preserved specimens and phylogenetic systematics of the genus Orestias (Teleostei: Cyprinodontidae). *Zootaxa*, *3640*(3), 373–394.

Hahn, E. E., Alexander, M. R., Grealy, A., Stiller, J., Gardiner, D. M., & Holleley, C. E. (2022). Unlocking inaccessible historical genomes preserved in formalin. *Molecular Ecology Resources*, *22*(6), 2130–2147.

Harrison, I., Abell, R., Darwall, W., Thieme, M. L., Tickner, D., & Timboe, I. (2018). The freshwater biodiversity crisis. *Science*, *362*(6421), 1369.

Hayes, M. M., & Armbruster, J. W. (2017). The Taxonomy and Relationships of the African Small Barbs (Cypriniformes: Cyprinidae). *Copeia*, *105*(2), 348–362.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003a). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321.

Hebert, P. D. N., Ratnasingham, S., & DeWaard, J. R. (2003b). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, *270*(SUPPL. 1), 96–99.

Heethoff, M. (2018). Cryptic Species – Conceptual or Terminological Chaos? A Response to Struck et al. *Trends in Ecology and Evolution*, *33*(5), 310.

Hortal, J., De Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 523–549.

Howes, G. J. (1987). The phylogenetic position of the Yugoslavian cyprinid fish genus Aulopyge Heckel, 1841, with an appraisal of the genus Berbus Cuvier & Cloquet, 1816 and subfamily Cyprininae. *Bulletin of the Britisch Museum of Natural History (Zoology)*, *52*, 165–196.

Howes, G. J. (1991). Systematics and biogeography: an overview. In I. J. Winfield & J. . Nelson (Eds.), *Cyprinid fishes: Systematics, biology and exploitation* (pp. 1–33). Chapman and Hall, London (UK).

Hubert, N., & Hanner, R. (2015). DNA Barcoding, species delineation and taxonomy: a historical perspective. *DNA Barcodes*, *3*(1), 44–58.

Hugueny, B., & Lévêque, C. (1994). Freshwater fish zoogeography in west Africa: faunal similarities between river basins. *Environmental Biology of Fishes*, *39*(4), 365–380.

Hykin, S. M., Bi, K., & McGuire, J. A. (2015). Fixing formalin: A method to recover genomic-scale DNA sequence data from formalin-fixed museum specimens using high-throughput sequencing. *PLoS ONE*, *10*(10), 1–16.

Kanu, U. C., Liang, C., Nwafor, C. C., Shen, J., & Zhang, E. (2022). Phylogenetic Analyses of Cyprinid Species from the Rokel. *Diversity*, *14*(4), 16. Retrieved from https://doi.org/10.3390/d14040299

King, J., & Brown, C. (2021). Africa's living rivers: Managing for sustainability. *Daedalus*, *150*(4), 240–259.

Kolbert, E. (2014). *Sixth extinction: an unnatural history* (Paperback.). London: Bloomsbury.

Leroy, B., Dias, M. S., Giraud, E., Hugueny, B., Jézéquel, C., Leprieur, F., Oberdorff, T., & Tedesco, P. A. (2019). Global biogeographical regions of freshwater fish species. *Journal of Biogeography*, *46*(11), 2407–2419.

Lévêque, C. (1983). Le genre Barbus (Pisces, Cyprinidae) en Côte d'Ivoire, *7*(3), 61–86.

Lévêque, C. (1994). Introduction générale: biodiversité des poissons africains. In G. G. Teugels, J.-F. Guegan, & J.-J. Albaret (Eds.), *Biological diversity of African Fresh- and Brackish water fiches.* (pp. 7–16).

Lévêque, C. (1997). *Biodiversity dynamics and conservation*. Cambridge University Press.

Lévêque, C. (2003). Cyprinidae. In D. Paugy, C. Lévêque, & G. G. Teugels (Eds.), *Faune des poissons d'eaux douces et saumâtres de l'Afrique de l'Ouest. Vol 1* (pp. 322–436). Paris.

Lévêque, C., Paugy, D., & Teugels, G. G. (1990). *Fresh and brackinsh water fishes of West Africa*.

Loiselle, P. V. (1973). Description of a new species of Barbus (Pisces Cyprinidae) from west Africa. Revue de Zoologie et de Botanique Africaines., *v. 87 (2)*, 379–390.

Lowe-McConnell, R. H. (1987). *Ecological studies in tropical fish communities*. Cambridge University.

Luttikhuizen, P. C., & Dekker, R. (2010). Pseudo-cryptic species Arenicola defodiens and Arenicola marina (Polychaeta: Arenicolidae) in Wadden Sea, North Sea and Skagerrak: Morphological and molecular variation. *Journal of Sea Research*, *63*(1), 17–23.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*(3), 523–536.

Maetens, H. (2019). The Enteromius species of the Lake Edward system A genetic and morphometric approach.

Maetens, H., Decru, E., Van Steenberge, M., Boom, A., & Snoeks, J. (n.d.). Uncovering the complex evolutionary history and species richness of the 'sawfin barbs' in the Lake Edward system (East Africa).

Maetens, H., Van Steenberge, M., Snoeks, J., & Decru, E. (2020). Revalidation of Enteromius alberti and presence of Enteromius cf. mimus (cypriniformes: Cyprinidae) in the lake edward system, East Africa. *European Journal of Taxonomy*, *2020*(700), 1–28.

Manda, B. K., Snoeks, J., Decru, E., Bills, R., & Vreven, E. (2020). Enteromius thespesios (Teleostei: Cyprinidae): a new minnow species with a remarkable sexual dimorphism from the south-eastern part of the Upper Congo River. *Journal of Fish Biology*, *96*(5), 1160–1175.

Manly, B. F. J. (1994). Multivariate statistical methods., *A primer,*(Boca Raton: Chapman & Hall).

Mullens, N., Sonet, G., Decru, E., Virgilio, M., Snoeks, J., & Vreven, E. (2020). Mitogenomic characterization and systematic placement of the Congo blind barb Caecobarbus geertsii (Cypriniformes: Cyprinidae). *International Journal of Biological Macromolecules*, *161*, 292–298.

Muschick, M., Nikolaeva, E., Rüber, L., & Matschiner, M. (2022). The mitochondrial genome of the red icefish (Channichthys rugosus) casts doubt on its species status. *Polar Biology*, *45*(10), 1541–1552.

Myers, G. S. (1960). Preface to any future classification of the cyprinid fishes of the genus Barbus. *Standford Ichtyol. Bull.*, *7*, 212–215.

National Academy of Sciences. (1988). *Biodiversity*. (E. O. Wilson, Ed.)*The National Academies Press*. Washington, DC. Retrieved from https://doi.org/10.17226/989

Nelson, J. S., Grande, T. C., & Wilson, M. V. H. (2016). Fishes of the world. *Fifth Edition*, 707.

Oh, J. H., Kim, S., & Lee, S. (2022). DNA barcodes reveal population-dependent cryptic diversity and various cases of sympatry of Korean leptonetid spiders (Araneae: Leptonetidae). *Scientific Reports*, *12*(1), 1–16.

Paireder, S., Werner, B., Bailer, J., Werther, W., Schmid, E., Patzak, B., & Cichna-Markl, M. (2013). Comparison of protocols for DNA extraction from long-term preserved formalin fixed tissues. *Analytical Biochemistry*, *439*(2), 152–160.

Pfingstl, T., Lienhard, A., Baumann, J., & Koblmüller, S. (2021). A taxonomist's nightmare – Cryptic diversity in Caribbean intertidal arthropods (Arachnida, Acari, Oribatida). *Molecular Phylogenetics and Evolution*, *163*(February), 107240.

Pikor, L. A., Enfield, K. S. S., Cameron, H., & Lam, W. L. (2011). DNA extraction from paraffin embedded material for genetic and epigenetic analyses. *Journal of Visualized Experiments*, (49), 15–17.

Piñeros, V. J., del R. Pedraza-Marrón, C., Betancourt-Resendes, I., Calderón-Cortés, N., Betancur-R, R., & Domínguez-Domínguez, O. (2022). Genome-wide species delimitation analyses of a silverside fish species complex in central Mexico indicate taxonomic over-splitting. *BMC Ecology and Evolution*, *22*(1), 1–19.

Poulin, R., & Pérez-Ponce de León, G. (2017). Global analysis reveals that cryptic diversity is linked with habitat but not mode of life. *Journal of Evolutionary Biology*, *30*(3), 641–649.

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery

for primary species delimitation. *Molecular Ecology*, *21*(8), 1864–1877.

QIAGEN. (2020). QIAamp ® DNA FFPE, (February). Retrieved from file:///C:/Users/Lenovo/Downloads/HB-0353-004_HB_QA_DNA_FFPE_0220_WW.pdf

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology and Evolution*, *36*(11), 1049–1060.

Reid, A. J., Carlson, A. K., Creed, I. F., Eliason, E. J., Gell, P. A., Johnson, P. T. J., Kidd, K. A., MacCormack, T. J., Olden, J. D., Ormerod, S. J., Smol, J. P., Taylor, W. W., Tockner, K., Vermaire, J. C., Dudgeon, D., & Cooke, S. J. (2019). Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological Reviews*, *94*(3), 849–873.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution*, 223–225.

Roberts, T. R. (1975). Geographical distribution of African freshwater fishes. *Zoological Journal of the Linnean Society*, *57*(4), 249–319.

Ruane, S., & Austin, C. C. (2017). Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. *Molecular Ecology Resources*, *17*(5), 1003–1008.

Schander, C., & Halanych, K. M. (2003). DNA, PCR and formalinized animal tissue - A short review and protocols. *Organisms Diversity and Evolution*, *3*(3), 195–205.

Schedel, F. D. B., Musilova, Z., Indermaur, A., Bitja-Nyom, A. R., Salzburger, W., & Schliewen, U. K. (2022). Towards the phylogenetic placement of the enigmatic African genus Prolabeops Schultz, 1941. *Journal of Fish Biology*, *101*(5), 1333–1342.

Schmidt, R. C. (2014). *Historical biogeography of fishes of the Fouta Djallon highlands and surrounding areas.* Doctoral dissertation, Tulane University School of Science and Engineering.

Schmidt, R. C., Bart, H. L., & Nyingi, W. D. (2017). Multi-locus phylogeny reveals instances of mitochondrial introgression and unrecognized diversity in Kenyan barbs (Cyprininae: Smiliogastrini). *Molecular Phylogenetics and Evolution*, *111*, 35–43.

Schmidt, R. C., Dillon, M. N., Kuhn, N. M., Bart, H. L., & Pezold, F. (2019). Unrecognized and imperilled diversity in an endemic barb (Smiliogastrini, Enteromius) from the Fouta Djallon highlands. *Zoologica Scripta*, *48*(5), 605–613.

Shokralla, S., Hellberg, R. S., Handy, S. M., King, I., & Hajibabaei, M. (2015). A DNA Mini-Barcoding System for Authentication of Processed Fish Products. *Scientific Reports*, *5*, 1–11.

Skelton, P. H., Tweddle, D., & Jackson, P. B. N. (1991). Cyprinids of Africa. In *Cyprinid fishes: Systematics, biology and exploitation* (pp. 211–239).

Snoeks, J. (2004). The Cichlid diversity of Lake Malawi/Nyasa/Niassa: identification, distribution and taxonomy., *El Paso*, *U*.

Snoeks, J., Harrison, I. J., & Stiassny, M. L. J. (2011). The status and distribution of freshwater fishes. In W. Darwall, K. Smith, D. Allen, H. Robert, I. Harrison, & E. Brooks (Eds.), *The Diversity of Life in African Freshwaters: underwater, under threat* (pp. 44–91). Gland, Switserland and Cambridge, UK: IUCN.

Straub, S. C. K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., & Liston, A. (2012). Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany*, *99*(2), 349–364.

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., Kistenich, S., Larsson, K. H., Liow, L. H., Nowak, M. D., Stedje, B., Bachmann, L., & Dimitrov, D. (2018). Finding Evolutionary Processes Hidden in Cryptic Species. *Trends in Ecology and Evolution*, *33*(3), 153–163.

Tang, C. Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T. G., & Fontaneto, D. (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(40), 16208–16212.

Teugels, G. G., & Hopkins, C. D. (2007). *The fresh and Brackish Water Fishes of Lower Guinea, West-Central Africa*.

Thieme, M. L., Abell, R., Stiassny, M. L. J., Skelton, P., & et al. (2006). *Freshwater ecoregions of Africa and Madagascar: a conservation assessment*. *Island Press* (Vol. 43).

United Nations. (1992). Convention on Biological Diversity, 1–28.

Van Ginneken, M., Decru, E., Verheyen, E., & Snoeks, J. (2017). Morphometry and DNA barcoding reveal cryptic diversity in the genus Enteromius (Cypriniformes: Cyprinidae) from the Congo basin, Africa. *European Journal of Taxonomy*, *2017*(310), 1–32.

Walters, A. D., Cannizzaro, A. G., Trujillo, D. A., & Berg, D. J. (2021). Addressing the Linnean shortfall in a cryptic species complex. *Zoological Journal of the Linnean Society*, *192*(2), 277–305.

Wheeler, Q. D. (2004). Taxonomic triage and the poverty of phylogeny. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1444), 571–583.

WWF. (2021). *THE WORLD ' S FORGOTTEN FISHES*. *WWF International*.

WWF. (2022). *Living Planet Report 2022 - Bulding a nature-positive society*. *WWF International*. Gland, Switserland. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3067

Yang, L., Sado, T., Vincent Hirt, M., Pasco-Viel, E., Arunachalam, M., Li, J., Wang, X., Freyhof, J., Saitoh, K., Simons, A. M., Miya, M., He, S., & Mayden, R. L. (2015). Phylogeny and polyploidy: Resolving the classification of cyprinine fishes (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution*, *85*, 97–116.

Zamani, A., Fric, Z. F., Gante, H. F., Hopkins, T., Orfinger, A. B., Scherz, M. D., Bartoňová, A. S., & Pos, D. D. (2021). DNA barcodes on their own are not enough to describe a species. *Systematic Entomology*, *47*(3), 385–389.