Population genomics and phylogenomics of two African freshwater sardines in a fisheries context

Leona Johanna Michèle Milec

A thesis for the degree of
Philosophiae Doctor (PhD)

PhD in Biosciences no. 67 (2024)
Faculty of Biosciences and Aquaculture

Doctor of Science: Biology (2024) - D/2024/2451/101
Faculty of Sciences, Hasselt University

PhD in Biosciences no. 67 (2024)
Leona Johanna Michèle Milec
Population genomics and phylogenomics of two African freshwater sardines in a fisheries context

© Leona Johanna Michèle Milec
978-82-93165-66-8

Print: Trykkeriet NORD

Nord University
N-8049 Bodø
Tel: +47 75 51 72 00
www.nord.no

All rights reserved.

No part of this book may be reproduced, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying or otherwise, without the prior written permission from Nord University.

Preface

This doctorate thesis is submitted in fulfilment of the requirements for the degree of Philosophiae Doctor (PhD) at the Faculty of Bioscience and Aquaculture (FBA), Nord University, Bodø, Norway, and the degree of Doctor of Science: Biology (2024) at the Faculty of Sciences, Hasselt University (joint PhD). The studies included in this dissertation represent original research as a part of the Stipendiat program at Nord University and the PhD program at Hasselt University.

The PhD project team consists of the following members:

Leona Johanna Michèle Milec, MSc: PhD candidate

Joost André Maria Raeymaekers, Professor, FBA, Nord University: primary supervisor

Maarten Pieterjan Maria Vanhove, Professor, Centre for Environmental Sciences, Research Group Zoology: Biodiversity & Toxicology, Hasselt University, Belgium: co-supervisor and primary supervisor at Hasselt University

Jorge Fernandes, Professor, FBA, Nord University: co-supervisor

Catherine Wagner, Associate professor, Department of Botany, University of Wyoming, USA: co-supervisor

Leona Johanna Michèle Milec

15th September 2024

Acknowledgements

The decision to take the leap and start this PhD was conscious, but nothing could have prepared me for the life-changing ride it turned out to be. It has changed me fundamentally as a person and taught me countless scientific and life lessons, both intended and unexpected. Most importantly, I have met some of the most amazing people that will be in my life forever. This section goes out to them and everyone else who has helped me and left their impressions on my memory of this PhD.

First and foremost, I would like to express my deepest gratitude to my two supervisors, Joost and Maarten. I will use a lot of fancy words but it will not be enough to express how thankful I am to them for passing on their passion and knowledge to me, pushing me when I needed to be pushed, and being patient when I needed patience. But most importantly, for always treating me as a colleague and friend more than a student, for always believing in me, and for their guidance, understanding, and encouragement when I was at my absolute lowest. I consider myself extremely lucky and would not have made it without two such wonderful supervisors who also turned out to be great friends.

Second, I would like to thank my co-supervisors Jorge Fernandes and Catherine Wagner for sharing their knowledge, insightful discussion and practical help with my research.

Third, I would like to thank all my fellow PhDs who have supported each other unconditionally during our time at Nord. Being in this together does really form a special kind of bond that I have not found anywhere else, nor can I properly explain it to anyone who was not a part of it. I will forever be grateful for the lovely community I found in Bodø. There are too many to name, but my time here was made unforgettable by all our dinners, strikk og drikks, parties, hiking, camping, mushroom picking, long nights in Sku Bare, campfires, swims, and climbing adventures. Above everyone else I would like to thank Amalia for being my wonderful partner in crime, a beautiful soul and the most hilarious, creative, and loving friend I have made in these

six years. My deep gratitude also goes out to the Bobbies Isabelle, Eric, Deepti and Valentin (and again Amalia) for being my family away from family, laughing and crying together and supporting each other through the toughest times, and to the BHALT! writing club Betty, Hirono, Thijs, (and again… Amalia) for motivating each other through the last parts of our PhDs even though we were far apart. Special thanks also goes out to Michael and Yousri for being excellent travel and adventure buddies and important parts of my mental support system, and to Solveig and Halvor, who taught me how to climb, as well as most of what I know about life in Norway, made us feel at home and helped us forget about all our troubles in their beautiful mountain cabin! Lastly, even though our ways have parted, I could not submit this thesis without thanking Alex for our countless laughs, for being so loving and supportive and for standing by me during the darkest period of my PhD.

Fourth, I would like to thank the whole LGG team for their fresh ideas and useful discussions, but also for combining work and fun with outdoor meetings and lab retreats, and generally providing an inspiring environment. I also want to thank Martina, Irina, Lars Martin, Jorge, Kostas, Anja, Brijesh, Srinidhi, Mathieu and Bart for lots of useful discussions and comments, and for their patient and enthusiastic help in the lab and with bioinformatics. Next, I want to thank the LBEG team at KUL for making my time in Leuven such a pleasant experience, and especially Els for our fun weeks in the lab working and learning together! My heartfelt thanks also goes out to our field team in Congo and our Congolese colleagues, in particular Pascal, Mulimbwa, Fidel, and Kaps, for a most interesting and inspiring collaboration, for making us feel at home, and for guiding us in their wonderful surroundings. Special thanks also goes to Maarten Van Steenberge for helping me with analyses and writing after my first rejection and beyond, and to my students, Lindah and Rita, for contributing passionately to this work and being a part of my learning process as much as I was to theirs. I'd like to thank the whole AB12 crew and parasitology-policy team at UHasselt, for another inspiring and encouraging environment, and for the warmest welcome I could wish for at a moment of my life when I really needed a warm welcome (and long nights at Fitlink). My

gratitude also goes out to the PXL BIO-Research team for a fun and stimulating working environment, and particularly Alain and Sarah for giving me the support and space I needed in my last year to finalize my thesis. During my last three PhD years, I received immense support from my friends in Belgium; especially from the Birdies and friends back home, LJBA (our consistent Tunezdays and LJBA radios provided me with the necessary soundtrack), the Radium mission crew, Team Endolphin, and the Gluglus for which I will be forever grateful. You gave me the energy I needed to finish this thesis!

Last and most, I would like to thank my family; my brother, my grandparents, and above all my parents for supporting me in every possible way, believing in me, and making me feel that they are proud of me from the start. I love you so much and without your love and support I could not even have started this journey, let alone finished it.

Contents

Preface I
Acknowledgements III
Contents VII
Abstract IX
Summary XI
Sammendrag på norsk XIII
List of figures and tables XV
List of abbreviations XVII
List of articles XIX
1 Introduction 1
1.1 Genetic and genomic tools in conservation and fisheries management 2
1.1.1 Stock identification 4
1.1.2 Genetic diversity 5
1.1.3 Effective population size 6
1.1.4 DNA-barcoding 7
1.1.5 Tracking introductions and invasions 8
1.1.6 Stakeholder challenges 9
1.2 Evolution of small pelagic fishes 11
1.2.1 General characteristics and ecology 11
1.2.2 Population structure 12
1.2.3 Speciation 14
1.3 Sex chromosome evolution and diversity in fishes 15
1.4 The African Great Lakes: a study system for fundamental and applied fish research 19
1.4.1 Geological history and diversification of fish lineages of the African Great Lakes 19
1.4.2 Small pelagic fishes in the African Great Lakes 20
1.4.3 The Lake Tanganyika sardines 25
1.4.4 A brief history of fisheries management in Lake Tanganyika 27
2 Overarching goal & aims 31
3 Main findings and discussion 33
3.1 New genomic resources for the Tanganyika sardines (Paper I & II) 33
3.2 Biogeographic context: diversification across Africa and intralacustrine speciation (Paper I) 37
3.3 Genomic and ecological factors underlying the divergence and contemporary evolution of the Tanganyika sardines (Paper I, II & III) 39

3.3.1 Factors contributing to between-species divergence: ecological opportunity, partial allopatry, and sex chromosome turnover 39
3.3.2 Factors contributing to within-species divergence: migration, inversion-facilitated selection and (historical) allopatry......................... 41
3.4 Recommendations for fisheries management of the Tanganyika sardines and elsewhere based on two genomic studies and one social study (Paper II, II & IV) 43
3.4.1 Genetic diversity and effective population size 43
3.4.2 Genetic population structure 45
3.4.3 Potential for implementation... 46
4 Conclusion 49
References 51

Abstract

Genetic diversity forms the basis of all biodiversity, and is increasingly recognized as an important component in the conservation and management of natural resources, including fisheries. Small pelagic fishes constitute nearly one third of global fish catches and are a critical food source, especially for low- and middle-income countries, and play key roles in aquatic ecosystems. Understanding their evolutionary history and contemporary genetic characteristics of these fish stocks is an important step in managing them sustainably. Small pelagic fishes have long been considered as panmictic with high genetic diversity due to limited barriers to gene flow. However, advancements in genomic technology have uncovered previously hidden population structure linked to genetic inversions and sex chromosomes, which can drive reproductive isolation and speciation even in seemingly well-mixed populations. Despite the growing availability of genomic data, its integration into fisheries management remains limited, largely due to communication gaps between stakeholders. Bridging these gaps requires understanding the diverse perspectives as well as regional variations in the acceptance of genomic applications in fisheries. Lake Tanganyika, the world's second-largest freshwater lake, faces significant threats from climate change and anthropogenic pressures, including overfishing and biodiversity loss. The lake's fisheries, dominated by two small pelagic species, *Stolothrissa tanganicae* and *Limnothrissa miodon*, are crucial for the food security of millions of people, but questions remain about their genetic diversity, population structure, and evolutionary history. In this thesis, we generated high-quality genomic resources for both species, revealing that historical allopatry, ecological opportunities, and sex chromosome turnover have been the key factors in their speciation. Within species, *S. tanganicae* exhibits near-panmictic structure due to gene flow, while *L. miodon* shows stronger population structure, both within and outside of Lake Tanganyika, driven by inversion-linked divergence in genes related to vision and sperm properties. We estimated policy-relevant genetic indicators, showing consistently high genetic diversity and effective population size across populations. Using a survey of African and

non-African fish and fisheries experts, we put our findings into a social and policy context, highlighting the challenges and opportunities for implementing genetic indicators in African fisheries management. Taken together, this thesis provides essential resources for the future genetic monitoring of Tanganyika sardines and offer concrete recommendations for the management of African fisheries.

Summary

Genetic diversity has become an unmissable staple of conservation and management of fish and fisheries. Measures of genetic diversity and its derived indices (e.g. population structure and connectivity, signatures of selection) provide information about the health of a fish stock and can aid fisheries managers in monitoring and decision making. Small pelagic fishes make up almost 30 % of global fish catches, provide abundant, cost-effective nutrition for low- and middle-income countries, and fulfil important roles in both marine and lacustrine ecosystems. It is therefore essential to gain knowledge on the evolutionary history and contemporary genetic properties of small pelagic fish stocks to ensure the continued sustainability of this valuable resource. Populations of small pelagic fishes, and marine fishes in general, have long been thought of as panmictic and highly genetically diverse due to a lack of barriers and high gene flow. However, advances in genomic technology have enabled detection of previously hidden and biologically relevant structure, often linked to structural rearrangements such as inversions, in some cases even leading to speciation. Another important driving force in the divergence and speciation of fishes is the evolution of new sex determination systems. These two powerful suppressors of recombination, inversions and sex chromosome turnover, have been able to cryptically generate partial or full reproductive isolation in populations and species that otherwise appear to be mixing without restriction. These mechanisms have thus received a growing amount of attention in recent nature conservation and fisheries genomics literature. Even though more and more genomic information continuously becomes available, most fisheries management schemes still do not incorporate it. This can mostly be attributed to communication and attitude differences between fisheries scientists, geneticists, policy makers and managers. Getting to know the different stakeholders, their backgrounds and opinions regarding the application of genomics in fisheries is the first step to closing this gap.

Lake Tanganyika, the world's second largest freshwater lake, is threatened by climate change and anthropogenic impacts under a rapidly growing population, including overfishing, deterioration of water quality, and a rapid loss of biodiversity. Two small pelagic fish species dominate the fisheries of Lake Tanganyika; *Stolothrissa tanganicae* and *Limnothrissa miodon*. These species feed millions of people, but appear to be suffering from anthropogenic impacts on the lake. Through this thesis, we aimed to further the genomic knowledge on the Lake Tanganyika sardines, shed light on ecological and genomic factors shaping their historical and contemporary evolution, and put these findings in a fisheries context. In **Paper I**, we sequenced their full mitochondrial genomes and used them to reconstruct the phylogeny of the tribe Pellonulini and more broadly the order Clupeiformes, to which the sardines belong. We estimated their divergence time and placed them into the geological context of Lake Tanganyika. In **Paper II**, we went further and sequenced and annotated the complete nuclear genomes of both sardines. We mapped lake-wide reduced-representation sequencing data to the genomes to characterize regions of elevated divergence between the sardines (sex chromosomes) and within *L. miodon* (large inversion). **Paper III** continues to apply these resources to introduced populations of *L. miodon* outside of Lake Tanganyika; those of Lake Kivu, Itezhi-Tezhi, Kariba, and Cahora Bassa, to quantify the genetic effects of their introduction. **Paper IV** then leaves the specific context of the Tanganyika sardines, and enters a social science context. In this chapter we address the lack of documentation of diverse opinions on the integration of genetic indicators into fisheries management. Specifically, we compare the opinions of African versus non-African fish and fisheries experts. Taken together, the four chapters of this thesis provide valuable resources for future genetic monitoring of the Tanganyika sardines and related pellonulines, a model system for the study of sex- and inversion-linked variation in native and introduced pelagic fish populations, and a social context for the practical application of genetic indicators in African countries.

Sammendrag på norsk

Genetisk mangfold har blitt et uunnværlig element i bevaring og forvaltning av fisk og fiskerier. Målinger av genetisk mangfold og tilhørende indekser (f.eks. populasjonsstruktur og konnektivitet, tegn på seleksjon) gir informasjon om tilstanden til en fiskebestand og kan hjelpe fiskeriforvaltere med overvåking og beslutningstaking. Små pelagiske fisk utgjør nesten 30 % av verdens fiskefangster, som gir rikelig og kostnadseffektiv ernæring til lav- og mellominntektsland og spiller en viktig rolle i både marine og lakustrine økosystemer. Det er derfor viktig å få kunnskap om evolusjonshistorien og de genetiske egenskapene til små pelagiske fiskebestander i dag for å sikre at denne verdifulle ressursen forblir bærekraftig. Populasjoner av små pelagiske fisk, og marine fisk generelt, har lenge vært ansett som panmiktiske og svært genetisk mangfoldige på grunn av mangel på barrierer og høy genflyt. Fremskritt innen genomteknologi har imidlertid gjort det mulig å oppdage tidligere skjulte og biologisk relevante strukturer, ofte knyttet til strukturelle omorganiseringer som inversjoner, som i noen tilfeller til og med har ført til artsdannelse. En annen viktig drivkraft i divergensen og artsdannelsen hos fisk er utviklingen av nye kjønnsbestemmende systemer. Disse to kraftige undertrykkerne av rekombinasjon, inversjoner og kjønnskromosomomsetning, har på kryptisk vis kunnet skape delvis eller full reproduktiv isolasjon i populasjoner og arter som ellers ser ut til å blande seg uhindret. Disse mekanismene har derfor fått stadig større oppmerksomhet i nyere litteratur om naturvern og fiskerigenomikk. Selv om det stadig blir mer genomisk informasjon tilgjengelig, tar de fleste fiskeriforvaltningsplaner fortsatt ikke hensyn til dette. Dette kan i stor grad tilskrives kommunikasjons- og holdningsforskjeller mellom fiskeriforskere, genetikere, beslutningstakere og forvaltere. Det første skrittet mot å tette dette gapet er å bli kjent med de ulike interessentene, deres bakgrunn og synet på anvendelsen av genomikk i fiskeriene.

Tanganyikasjøen, verdens nest største ferskvannssjø, er truet av klimaendringer og menneskeskapte påvirkninger som følge av en raskt voksende befolkning, inkludert

overfiske, forringelse av vannkvaliteten og et raskt tap av biologisk mangfold. To små pelagiske fiskearter dominerer fiskeriene i Tanganyikasjøen: *Stolothrissa tanganicae* og *Limnothrissa miodon*. Disse artene gir mat til millioner av mennesker, men ser ut til å lide under den menneskeskapte påvirkningen på innsjøen. Gjennom denne avhandlingen ønsket vi å øke den genomiske kunnskapen om sardiner i Tanganyikasjøen, belyse økologiske og genomiske faktorer som har formet deres historiske og moderne utvikling, og sette disse funnene inn i en fiskerikontekst. I **Artikkel I** sekvenserte vi deres komplette mitokondriegenomer og brukte dem til å rekonstruere fylogenien til stammen Pellonulini og mer generelt til ordenen Clupeiformes, som sardinene tilhører. Vi har estimert tiden for deres divergens og satt dem inn i Tanganyikasjøens geologiske kontekst. I **Artikkel II** gikk vi videre og sekvenserte og annoterte de komplette kjernegenomene til begge sardinene. Vi kartla sekvenseringsdata med redusert representasjon for hele innsjøen til genomene for å karakterisere regioner med høy divergens mellom sardinene (kjønnskromosomer) og innad i *L. miodon* (stor inversjon). I **Artikkel III** fortsetter vi å anvende disse ressursene på introduserte bestander av *L. miodon* utenfor Tanganyikasjøen, nemlig i Kivusjøen, Itezhi-Tezhi, Kariba og Cahora Bassa, for å kvantifisere de genetiske effektene av introduksjonen. **Artikkel IV** forlater deretter den spesifikke konteksten til Tanganyika-sardinene og går inn i en samfunnsvitenskapelig kontekst. I dette kapittelet tar vi for oss mangelen på dokumentasjon av ulike meninger om integrering av genetiske indikatorer i fiskeriforvaltningen. Vi sammenligner spesielt oppfatningene til afrikanske og ikke-afrikanske fiske- og fiskerieksperter. Til sammen gir de fire kapitlene i denne avhandlingen verdifulle ressurser for fremtidig genetisk overvåking av Tanganyika-sardiner og beslektede pellonuliner, et modellsystem for studier av kjønns- og inversjonsrelatert variasjon i stedegne og introduserte pelagiske fiskebestander, og en sosial kontekst for praktisk anvendelse av genetiske indikatorer i afrikanske land.

List of figures and tables

Figure 1. Number of publications per year of commercial fisheries research using high-throughput sequencing platforms. The pie chart shows different sequencing platforms used in these publications. Adapted from Valenzuela-Quiñoez (2016). 3
Figure 2. Map of the African Great Lakes region. From https://www.greatlakesofafrica.org/about-the-lakes/ 20
Figure 3. Photographs of the Lake Tanganyika sardines a) *Stolothrissa tanganicae* and b) *Limnothrissa miodon*, showing their morphological differences. Adapted from Abdulkarim et al. (2014) and Pekka Kotilainen. 26

Table 1. Overview of small pelagic fishes in the African Great Lakes. Local names and yield estimates are from Kolding et al. (2019) unless otherwise indicated. Habitat and distribution information are from FishBase (Froese and Pauly 2024). 22

List of abbreviations

CBD = Convention of Biological Diversity

COI = cytochrome *c* oxidase subunit 1

DNA = deoxyribonucleic acid

eDNA = environmental DNA

FAO = Food and Agriculture Organization of the United Nations

FISH-BOL = fish barcode of life project

IBD = isolation-by-distance

IUU = illegal, unreported, and unregulated fishing

LTA = Lake Tanganyika Authority

LTBP = Lake Tanganyika Biodiversity Project

LTR = Lake Tanganyika Research

LTRIMP = Lake Tanganyika Regional Integrated Management Programme

MY(A) = millions of years (ago)

N_c = effective population size

N_e = census population size

qPCR = quantitative PCR

RFLP = restriction fragment length polymorphisms

SNP = single nucleotide polymorphism

List of articles

Paper I

Leona J. M. Milec, Maarten P. M. Vanhove, Fidel Muterezi Bukinga, Els L. R. De Keyzer, Vercus Lumami Kapepula, Pascal Mulungula Masilya, N'Sibula Mulimbwa, Catherine E. Wagner and Joost A. M. Raeymaekers. Complete mitochondrial genomes and updated divergence time of the two freshwater clupeids endemic to Lake Tanganyika (Africa) suggest intralacustrine speciation (2022). *BMC Ecology & Evolution* 22(127), 1-7. https://doi.org/10.1186/s12862-022-02085-8

Paper II

Leona J. M. Milec, Catherine E. Wagner, Fidel Muterezi Bukinga, Els L.R. De Keyzer, Vercus Lumami Kapepula, Nikol Kmentová, Pascal Masilya Mulungula, N'Sibula Mulimbwa, Jessica A. Rick, Kostas Sagonas, Maarten Van Steenberge, Srinidhi Varadharajan, Brijesh S. Yadav, Maarten P.M. Vanhove, Filip A.M. Volckaert, Anja M. Westram and Joost A.M. Raeymaekers (2024). Sex chromosome turnover and inversion-linked adaptation in two African freshwater sardines (Manuscript).

Paper III

Leona J. M. Milec, Maarten P.M. Vanhove, Jorge M. O. Fernandes, Bart Hellemans, Cyprian Katongo, Jeppe Kolding, Nikol Kmentová, Pascal Masilya Mulungula, Claque J. Maunde, Tamuka Nhiwatiwa, Maarten Van Steenberge and Joost A.M. Raeymaekers (2024). Out of Tanganyika: genetic effects of the introduction and invasion of *Limnothrissa miodon* into natural and man-made African lakes (Manuscript).

Paper IV

Leona J. M. Milec, Jean Hugé, Joost A. M. Raeymaekers, Sophie Van Schoubroeck, Maarten Van Steenberge and Maarten P. M. Vanhove (2024). High expectation, low implementation: perceptions of African fish and fisheries experts on genetic indicators in fisheries management (Manuscript submitted to *Environmental and Sustainability Indicators*). Available at SSRN: https://ssrn.com/abstract=4917674

1 Introduction

Through a combination of genomic and social studies, this thesis explores the ecological and genomic factors contributing to evolutionary divergence between and within two African freshwater sardine species in a fisheries context. To establish the necessary methodological foundation, we first introduce the use of genetic and genomic tools in tackling contemporary challenges in conservation and natural resource management, specifically fisheries management. We discuss key tools and concepts; stock identification, genetic diversity, effective population size, DNA barcoding, and the tracking of introductions and invasions. Incorporation of these methods in fisheries management poses challenges for fisheries scientists, managers, and other stakeholders. These challenges vary between stakeholders of different geographical and demographic backgrounds, interests, and influence, which are especially diverse in the field of fisheries. Understanding the real-life applicability of genetic markers, indicators, and tools in different contexts require social science approaches.

Given that the study species in this thesis are small freshwater sardine-like fishes from the African Great Lakes region, we elaborate on small pelagic fishes and discuss their general characteristics, population structure, and speciation processes. We zoom in on two mechanisms which appear to play a disproportionate role in the divergence within and between various marine and freshwater fishes, including small pelagic species: structural chromosomal rearrangements (inversions), and the evolution of sex chromosomes. Small pelagic fishes are crucial to food security in the African Great Lakes region, with many species remaining understudied. We thus describe the fisheries and available genetic work on small pelagic fishes in the African Great Lakes. We conclude by introducing the sister species pair *Stolothrissa tanganicae* and *Limnothrissa miodon,* also known as the Tanganyika sardines, as a study system for both fundamental research on factors driving evolution of small pelagic fishes, and for the integration of new genomic knowledge in fisheries management.

1.1 Genetic and genomic tools in conservation and fisheries management

Genetic diversity forms the basis of all biodiversity. Within species, differences in DNA sequence between individuals arise through mutation, random mating, and gene flow. These differences are the substrate on which natural selection can act, allowing organisms to adapt to changes in their environment. Conservation biology, the study of protecting species from extinction and their ecosystems from loss of functioning, has been incorporating genetic diversity almost from the time of its conception. The Convention of Biological Diversity (CBD), the international legal instrument which unites 196 countries in their aim to conserve biodiversity, defines 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 part; this includes diversity within species, between species and of ecosystems" (CBD 2011). The ultimate goal is "the sustainable use of the components of biodiversity and the fair and equitable sharing of the benefits arising out of the utilization of genetic resources" and the inclusion of genetic aspects has been fortified in recent calls for action (CBD 2020; Hoban et al. 2020). Nowadays, conservation genetics is a mature field with direct application of theoretical concepts from population genetics (Willi et al., 2022). This has allowed conservation geneticists to closely follow developments in DNA-sequencing and analysis technology. In the last two decades, conservation genetics, which operates at the level of single genes or sites in the genome, has evolved into conservation genomics, which widens the view to whole genomes and dramatically increases the resolution at which populations can be studied (Primmer 2009; Allendorf et al. 2010; Ouborg et al. 2010; Theissinger et al. 2023).

The inherent value and beauty of nature can be reason enough for many people to justify species and ecosystem preservation (Chan et al. 2016), but ensuring provision of natural resources to humans is the major motivation in our anthropocentric world. Accordingly, global initiatives to conserve genetic diversity currently place emphasis on

domesticated or commercially important species (Laikre et al. 2020), and governments tend to fund conservation of species with economic value (Bakker et al. 2010). Different natural resources have seen the development of their own specialized fields of conservation genetics, such as hunting of game species (Bieniek-Kobuszewska et al. 2020; Chen et al. 2023), wild crop relatives (Wambugu and Henry 2022), forestry (Koskela et al. 2013), and fisheries (Allendorf et al. 1987; Bernatchez et al. 2017). The field of fisheries science caught on to genetics as a management tool between the 1970s and 1990s (Allendorf et al. 1987). In parallel to conservation genetics, it has readily adapted to the increasing complexity and inflow of genomic data in the last two decades (Bernatchez et al. 2017; Lu and Luo 2020) (Figure 1).

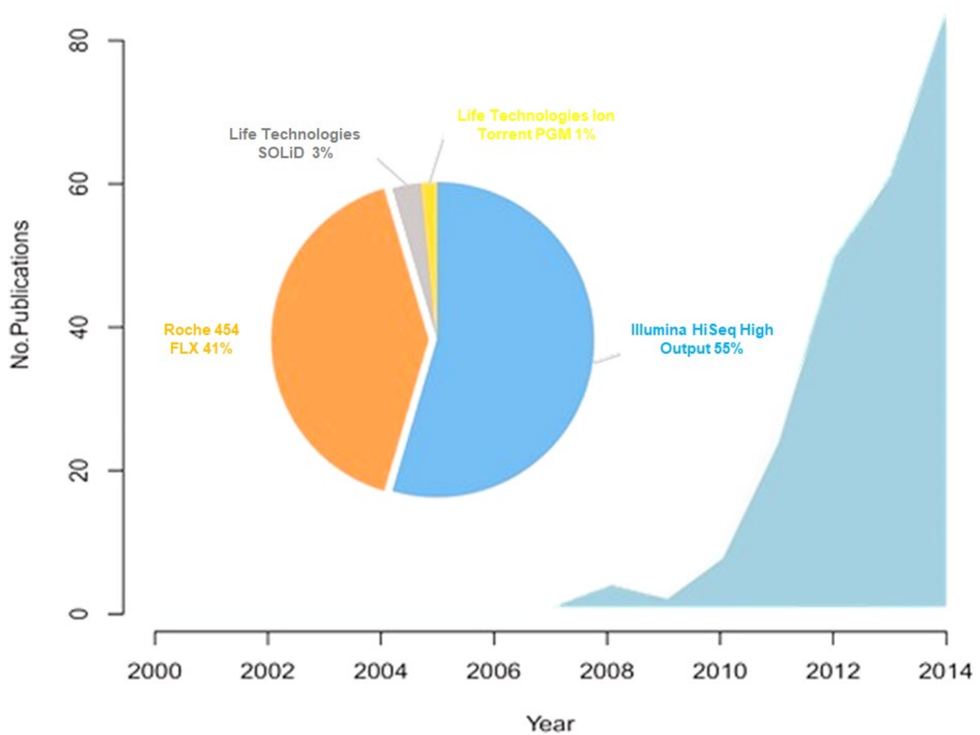

Figure 1. Number of publications per year of commercial fisheries research using high-throughput sequencing platforms. The pie chart shows different sequencing platforms used in these publications. Adapted from Valenzuela-Quiñoez (2016).

1.1.1 Stock identification

A population in biology is defined as a collection of individuals of the same species that inhabit the same geographical area and are capable of breeding with each other. This implies a degree of genetic distinctness from other populations and adaptation to local conditions (Primmer 2009; Ouborg et al. 2010). Populations can be delineated by comparing within-group with between-group genetic variation (i.e. many shared variants *versus* many variants private to each group). Delineation of populations is the most widely applied concept from population genetics in fisheries (Dichmont et al. 2012). In fisheries management, populations are commonly referred to as 'stocks', although definitions of a stock contain nuances specific to the field. In the broadest sense, a stock describes a group of individuals with characteristics that are considered homogenous for a particular management purpose (also called a 'management unit', Begg and Waldman, 1999), and can even encompass different species (European Environment Agency 2023). A variety of methods has been applied to define stocks, from mark-recapture of individuals, catch data, life history and behavioural observations, to morphometrics, the study of parasites and otoliths, and genetics (Begg and Waldman 1999).

The genetic definition of a stock, a local subpopulation of a single fish species with genetic differences from other subpopulations (MacLean and Evans 1981), most closely corresponds to the mechanism that structures populations in the strict sense. As population delineation, genetic stock identification is based on the distribution of genetic variation along the geographical range of the target species. High connectivity causes a homogenous distribution, while limits to dispersal cause more genetic differentiation, with completely separated stocks being highly genetically distinct (Shaklee et al. 1999; Dichmont et al. 2020). Disregard of genetic structure in management can damage stocks in two ways. If one panmictic or several well-connected populations are managed as separate ones, fishing quotas for the individual populations may add up to more than the entire stock can take, putting it at risk of overexploitation. If the separate fishing quotas are set lower than necessary, they can

lead to a loss of yield. Conversely, if (partially) separate subpopulations are managed as a single stock, falsely assuming that fish can be refilled from other areas, the main risk is local depletion (Waples et al. 2008; Ying et al. 2011).

1.1.2 Genetic diversity

Genetic diversity can be divided into two types; neutral and adaptive genetic diversity. Neutral genetic variants do not have consequences for an individual's fitness and are thus under no or very weak selection. Examples are synonymous amino acid substitutions or mutations at non-coding sites. Adaptive genetic variants have fitness consequences and are under negative, positive, or balancing selection by abiotic or biotic conditions (Holderegger et al. 2006). Adaptive genetic diversity, a proxy for evolvability or adaptive potential, is of great interest to conservation genetics, as it provides populations, species, and communities, with resilience in the face of environmental changes and exploitation (Schindler et al. 2015; Crutsinger 2016; Willi et al. 2022).

Fisheries can benefit from incorporating both measures of genetic diversity. Neutral genetic diversity can be used to estimate demographic parameters, such as effective population size (see section 1.1.3) and its changes over time, connectivity and dispersal, genetic drift, and neutral genetic differentiation (see section 1.1.1). Adaptive genetic diversity can help to resolve population structure, especially in highly mobile species, where neutral genetic divergence is often low. Locally selected genetic variants are usually more differentiated than neutral variants, and can thus provide a higher degree of certainty for assigning individuals to stocks (Hauser & Carvalho, 2008; Nielsen et al., 2009). In addition, adaptive genetic diversity has helped many fisheries target species, such as Atlantic cod and European hake, to adapt to changes in temperature and salinity (Hauser and Carvalho 2008; Valenzuela-Quiñoez 2016). Consequently, it is essential to preserve adaptive genetic variation across stocks, to provide fishes with the potential to cope with further climate change.

1.1.3 Effective population size

In conservation genetics, population size has two primary definitions; census population size (N_c) and effective population size (N_e). N_c is defined as the number of (adult or breeding) individuals in a population and can be estimated by simple counting, genetic and non-genetic mark-recapture techniques, or detection probability modelling (Luikart et al. 2010). N_e is defined as the number of effectively reproducing individuals, or those that actually pass on their genes to the next generation. More broadly, N_e measures the change in neutral genetic variation from one generation to the next as a result of random sampling of alleles (i.e. genetic drift) (Charlesworth 2009; Luikart et al. 2010). It can be estimated based on several sources of genetic data; the non-random association of alleles (linkage disequilibrium), relatedness between individuals, the excess of heterozygosity, or temporal differences in allele frequencies (Charlesworth 2009; Wang et al. 2016).

N_e is directly related to other parameters that are of interest for conservation (Frankham 2010). On evolutionary timescales, mutation rate (μ) and N_e together determine the amount of neutral genetic variation in a population, with larger ones generally being more diverse. Natural selection is also less effective, and the effects of inbreeding and genetic drift are stronger, in small populations (Kimura 1983; Charlesworth 2009). It follows that abrupt and/or large changes in population size and N_e, for example through overexploitation or introduction of a small founder population, reduce genetic diversity and increase the chances of inbreeding and genetic drift, an effect known as a genetic bottleneck (Allendorf and Lundquist 2003). Contrarily, population expansion usually increases genetic diversity, although this takes time and not all populations can recover after a harsh bottleneck (Charlesworth 2009).

In fisheries management, estimates of N_e can be used as a proxy of abundance (Luikart et al. 2010; Ovenden et al. 2016). They can provide an alternative or complement to catch data, which can sometimes be misleading, and labour-intensive time-series (e.g. scientific fishing, visual and acoustic surveys, mark-recapture). A meta-analysis and

simulation study of 32 overfished and 108 non-overfished species clearly showed that in general, overfishing-induced bottlenecks cause a reduction of genetic diversity (Pinsky and Palumbi 2014). Its direct link to genetic diversity and the strength of selective and drift effects makes N_e a valuable parameter to predict the evolutionary trajectories of exploited species.

1.1.4 DNA-barcoding

DNA-barcoding makes use of a species-specific or population-specific 'fingerprint' of a short DNA sequence for species determination of a (preserved) specimen, tissue fragment, or even an environmental sample such as soil or water (Valentini et al. 2009). While initially intended for sample identification, the DNA barcode framework has also been useful for species and population delineation and the description of cryptic diversity that could not be detected through traditional taxonomy (Hubert and Hanner 2015). For animals, the most commonly used marker is the mitochondrial gene cytochrome *c* oxidase subunit 1 (COI). However, the applicability of COI is not universal, as some species harbour higher intra- than interspecific variation at this marker (Vanhove et al. 2013; Hubert and Hanner 2015). Using other nuclear and mitochondrial genes is possible, but the applicability also differs between taxonomic groups (Zardoya and Meyer 1996; Duchêne et al. 2011; Kmentová et al. 2023). Barcoding has been extensively applied to fisheries, both for fundamental research and practical questions. The fish barcode of life project (FISH-BOL) and FishTrace are initiatives that aim to create barcode reference databases for all ~ 30,000 fish species (Ward et al. 2009), and European fisheries targets (Zanzi and Martinsohn 2017), respectively. In efforts to describe fish diversity, barcoding can complement or replace traditional taxonomic methods, which can be time-consuming and heavily rely on specialist expertise. In addition, small fragments, processed products, endangered species, and non-adult life stages (e.g. eggs, larvae, small fry) can be difficult to identify visually (Ward et al. 2009). Practical applications of barcoding in fisheries management include species and population-level identification of landings and by-catches, conservation of threatened species and highly diverse ecosystems, and tracking invasive alien species (Ward et al.

2009; Pavan-Kumar et al. 2016). In addition, it has helped in fighting accidental and intentional mislabelling of and mis-certification of fish products, illegal ornamental fish trade, and illegal, unreported, and unregulated (IUU) fishing (Ogden, 2008; Ward et al., 2009; Nielsen et al., 2012).

Standard COI-based barcoding is more difficult in clades where diagnostic mutations are lacking, where within-species variation is higher than between-species variation, where divergence is recent, and where hybridization occurs (Ward et al. 2009; Pavan-Kumar et al. 2016), the latter being exceptionally common in fishes (Ryman et al. 1995). In addition, this approach fully relies on high quality reference databases, which are often incomplete (Pereira et al. 2013; FAO 2019; Decru et al. 2022). FISH-BOL also aims to include several specimens per species to enable population-level assignment for species with strong spatial differentiation, and the detection of cryptic species (Ward et al. 2009). Barcoding studies often use phylogenetic trees to visualize divergence among samples. On the other hand, phylogenetic reconstructions can be the *per se* goal of barcode-based approaches. While mitochondrial genes such as COI can lead to robust phylogenetic hypotheses (Gaunt and Miles 2002; Brant and Loker 2009), in many cases there are strong drawbacks to using only mitochondrial markers, which are discussed in depth in **Paper I**.

1.1.5 Tracking introductions and invasions

For over a century, fishes have been introduced into non-native ecosystems for a variety of reasons, most often to improve recreational or artisanal fisheries. Accidental introductions are also common, be it through the natural connectivity of water bodies, the transport industry, or escapees from aquaculture farms and aquaria (Lévêque 1996; Rius et al. 2015; Bernery et al. 2022). While there can be commercial or recreational benefits for humans, many introductions/invasions have had negative consequences on local ecosystems (Ogutu-Ohwayo and Hecky 1991; Bernery et al. 2022). The establishment success of non-native species depends on the environmental conditions and stability in the new habitat, availability of an empty niche, the position of the

species in its new food web, its competitive ability, life history, and whether it is specialist or generalist (Miller 1989; Allendorf and Lundquist 2003; Olden et al. 2006; Bernery et al. 2022). Using the principles described in sections 1.1.1 – 1.1.3, the history of introduced or invasive species can be reconstructed by mapping genetic diversity and population structure, establishing relatedness between source populations and derived populations, and estimating current and past population sizes, i.e. populations expansions and contractions (Kamenova et al. 2017).

The size of the founder population is of particular interest in determining introduction or invasion success. First, the number of reproductively able individuals determines population growth. Second, classic population genetic theory predicts that a small founder population passes through a genetic bottleneck, causing loss of evolutionary potential and inbreeding depression. However, a recent review of introductions of marine fishes, invertebrates, and plants into Europe revealed that these effects are rather uncommon (Rius et al. 2015). The pervasive success of introduced populations, which are typically small and not locally adapted, thus becomes a paradox (Allendorf and Lundquist 2003). If a population grows rapidly after a founder event, reduction of genetic diversity can be limited (Frankham 2005). Additionally, multiple founding or release events from genetically diverse origin populations, besides the obvious effect of adding more individuals (increased 'propagule pressure'), can supplement the genetic diversity of the introduced population(Allendorf and Lundquist 2003; Frankham 2005; Rius et al. 2015; Bernery et al. 2022).

1.1.6 Stakeholder challenges

It is clear from the above examples that the field of fisheries genetics and genomics is well developed, with principles established and guidelines provided (Dichmont et al. 2012; Bernatchez et al. 2017). The question arises why many species are not managed accordingly (Waples et al. 2008; Reiss et al. 2009; Osio et al. 2015; Mullins et al. 2018). One reason is that biological borders or clines often do not overlap with political and economic ones, rendering genetically determined stock boundaries impractical

(Waples et al. 2008; Reiss et al. 2009). Non-genetic fisheries science is also a mature field, which is reflected in a tradition of non-genetic stock assessment methods. Furthermore, there is a lack of transparency, communication, and mutual understanding between geneticists, non-genetic fisheries scientists, managers, and other stakeholders (Mora et al. 2009; Bernatchez et al. 2017). Validating and translating specialized methods into cheap, easy-to-use tools and application to management is a challenging process with many steps (Ogden 2008; Waples et al. 2008). Lastly, while some improvement has been noted, perceptions of genetic methods among non-geneticists are still quite negative. Genetic methods are often regarded as expensive, time-consuming, training-intensive, lacking robustness or consistency, exhibiting too much uncertainty, and not always in line with management goals (Waples et al. 2008; Dichmont et al. 2012; Ovenden and Moore 2016).

In order to increase transparency and trust around the incorporation of genetic methods in fisheries management, it is necessary to understand the needs and expectations of all involved stakeholders. Fisheries stakeholders are diverse in their background, level of interest, and level of influence in a fishery. They can be private persons, members of (non-)governmental, regional and intergovernmental bodies, and private sector organizations, e.g. consumers, fishers, local traders, fisheries scientists, managers, activists, business owners, or company employees (Pomeroy et al., 2016; De Keyzer et al., 2020). Discrepancy of opinions is thus not surprising, and it is essential to incorporate expectations and needs of all players. A formal stakeholder analysis is the first step in this process. By mapping all potential stakeholders on the interest-influence continuum, policy-makers can understand whose interests they need to consider, what these interests are, what stakeholders can offer to the management process, and how they might react to certain policies (Brugha 2000). Surveys or face-to-face interviews with the stakeholders can be used to document the opinions, needs and expectations on how and why a fishery should be managed. Ideally, representatives of all stakeholder groups are not only included as study subjects, but

also actively participate in the policy-making process and the research feeding into these policies (Schwermer et al. 2018).

1.2 Evolution of small pelagic fishes

1.2.1 General characteristics and ecology

Small pelagic (open water) fishes are an essential food source for humans across the globe. They constitute 28% of global wild fish catches (MSC 2021) and provide the most abundant and cost-effective nutrition out of all fish species for low- and middle-income countries (Robinson et al. 2022). Small pelagic fishes also play a critical role in marine and freshwater food webs. By feeding on (zoo)plankton and being eaten by larger marine species, they form a funnel of energy between low and high trophic levels, a function that is often fulfilled by only one or a few small pelagic species (Cury 2000). Small pelagic fishes are taxonomically diverse, represented for example by members of Clupeiformes (herring, anchovies, sardines, …), Scombridae (mackerels), Carangidae (scads), Atherinidae (silversides), and many others (Stephenson and Smedbol 2001). Nevertheless, they share many of their behavioural and life history characteristics. In general, small pelagic fishes are highly mobile, migrate across large distances, form large schools and occupy wide geographical distributions (Stephenson and Smedbol 2001). They are r-selected ('live fast and die young') species that grow rapidly, reach maturity early, have high fecundity and a short lifespan (Adams 1980). Many small pelagic species inhabit upwelling systems, where seasonal variation in ocean currents results in massive bursts of nutrient supply, and consequently food (plankton) provision for these fish. This implies that they are subject to 'boom and bust' dynamics, i.e. strong fluctuations in abundance caused by bottom-up (climate and food availability), top-down (fishing and predation), and parallel (interspecific competition) forces, both on ecological and evolutionary timescales (Grant and Bowen 1998; Cury 2000; Peck et al. 2021). Considering their critical role in human food security and aquatic food webs, overfishing of small pelagic fishes is a major concern. Their schooling behaviour makes them particularly vulnerable to detection when fishers use

sonar to locate stocks, and to harvesting of large numbers at a time, while concealing reductions in overall stock size (Beverton 1990; Stephenson and Smedbol 2001). Despite their high abundance and naturally strong fluctuations, many collapses of small pelagic fish stocks have been documented, not always followed by recovery after reduction of fishing (Beverton 1990). While some small pelagic species such as the European sardine (*Sardina pilchardus*) are currently sustainably fished or even underfished, others such as round sardinella (*Sardinella aurita*) and the serra Spanish mackerel (*Scomberomorus brasiliensis*) are considered overfished (FAO 2022).

1.2.2 Population structure

In population genetic terms, high abundance and mobility in a barrier-poor landscape translates into large effective population sizes and a high amount of gene flow. Especially in the openness of the marine pelagic realm, it is tempting to think of such populations as homogenous across continental scales, highly diverse, and resilient to the loss of genetic diversity (Ryman et al. 1995; Hauser and Carvalho 2008). Indeed, population structure is often extremely shallow in marine fishes (Grant and Bowen 1998). Early genetic studies using electrophoretic markers overall showed high within-population and low between-population differentiation in marine compared to freshwater fishes, presumably due to a lack of barriers and high mobility at different life stages (Ward et al. 1994). However, a small number of genetic differences may have large biological significance and selection is more effective in large populations, favouring the emergence of locally adapted alleles even in the face of gene flow (Kimura 1983; Charlesworth 2009). Large populations may also be especially sensitive to long-term effects of population contractions because they proportionally lose more alleles than smaller populations (Ryman et al. 1995). While more studies have focused on these principles in marine fishes, the same principles hold up for freshwater small pelagic fishes, which are also numerous (Lavoué et al. 2013; Kolding et al. 2019).

Despite the apparent lack of barriers and population structure, genetic differentiation could be identified in some small pelagic species. Structuring factors include

bathymetric constraints such as straits and basin borders (Hauser and Carvalho 2008), fronts between distinct water bodies (Vandamme et al. 2021), environmental gradients such as temperature, salinity, and nitrate concentration (Barrio et al. 2016; Canales-Aguirre et al. 2016), regional current patterns and coastal structure (Glover et al. 2011; Canales-Aguirre et al. 2016), behavioural adaptations such as sea-going *versus* resident behaviour (Glover et al. 2011), spring- *versus* autumn-spawning (Petrou et al. 2021), and homing behaviour (Ruzzante et al. 2006).

As the cost of sequencing technology dropped during the last few decades and better genomic resources became available, more previously hidden structure was revealed for many commercially important small pelagic species. Thanks to the improved resolution, genetic variation at adaptive loci could be identified and linked to biological functions (Hauser and Carvalho 2008; Bernatchez et al. 2017). For example, in Atlantic herring, whole genome sequencing revealed large blocks of loci involved in differentiating spring- and autumn-spawning populations regardless of geographic origin. The same study found hundreds of loci at the basis of adaptation to lower salinity levels following a range expansion into the Baltic Sea (Barrio et al. 2016). Shortly after, a chromosome-level assembly traced the spring-autumn spawning variation back to a large structural rearrangement (chromosomal inversion) (Pettersson et al. 2019). Similar inversion-linked differentiation according to spawning phenology was found in Pacific herring, with a fine-scale north-south gradient in the order of weeks to months (*C. pallasii*) (Petrou et al. 2021). These recent studies in pelagic fishes attributed genetic structure among spatially homogenous populations to chromosomal regions of reduced recombination. These regions are less exposed to the homogenizing effects of gene flow (Turner et al. 2005; Yeaman and Whitlock 2011; Jones et al. 2012) and can be created by chromosomal rearrangements such as inversions (Kirkpatrick and Barton 2006), duplications (Yeaman and Whitlock 2011) or hitchhiking with adaptive loci (Via 2012). Association of divergent loci is thought to be the result of selection favouring certain combinations of genes or traits and preventing recombination of these complexes (Hill and Zhang 2012; Via 2012). In situations of high

gene flow, such regions can be the key to local adaptation or even speciation (Kirkpatrick and Barton 2006; Hoffmann and Rieseberg 2008). Inversions are more likely to contribute to divergence between incipient species if they are large, and if gene flow is high compared to selection strength per locus (meaning the inversion captures many small-effect loci) (Feder et al. 2014).

1.2.3 Speciation

The same processes that cause intraspecific differentiation can drive speciation through continued allopatric divergence (e.g. by distance, ecological or geographic barriers). Alternatively, for sympatric divergence to occur, there needs to be selection against hybrids between ecologically diverging populations ('reinforcement'), coupled with a mechanism of reproductive isolation (Rundle & Nosil, 2005). The mechanisms that cause speciation in small pelagic fishes imply allopatric speciation as the most common mode (Lavoué et al. 2013; Gaither et al. 2016; Silva et al. 2017). Aggregation at specific sites outside of the pelagic realm, for example for breeding, offer opportunities for within-population divergence (Ruzzante et al. 2006), and potentially also for speciation, which can still be considered allopatric (Fryer 2006).

Speciation in small pelagic fishes has mostly been studied on a macroevolutionary scale. Within clades of small pelagic fishes (e.g. clupeiforms), the large number of habitat transitions, for example between pelagic, littoral, demersal, or riverine environments, yields higher rates of speciation than within the pelagic realm (Fryer 2006; Bloom and Egan 2018; Bloom and Lovejoy 2014; Egan et al. 2018). Several studies have addressed the low level of interspecific compared to intraspecific divergence within pelagic clupeoid genera that traditionally consist of multiple species, including *Sardinella* (Stern et al., 2018), old world anchovies of *Engraulis* (Grant and Bowen, 1998; Silva et al., 2017), *Sardina* and *Sardinops* (Grant and Bowen 1998). These studies all question the validity of some of the species separations and propose the concept of cosmopolitan species with local phenotypic variation, or incipient speciation. Similar patterns have been found in larger circumtropical pelagic species (Gaither et al., 2016).

The main driving forces for this shallow 'interspecific' differentiation are vicariance and oceanographic boundaries, climate oscillations causing periodic contraction and expansion or even local extinctions and recolonizations, and contemporary gene flow. These processes can play out at regional or continental scales (Grant and Bowen 1998). High mobility at the adult stage, as well as the ability to tolerate a wide range of environmental conditions, are crucial for a species to reach global distribution and at the same time prevent speciation even on a global scale (Gaither et al. 2016).

1.3 Sex chromosome evolution and diversity in fishes

Our stereotypical view of sex chromosomes is based on humans; a fixed female (XX) – male (XY) system that has been stable across all mammals for millions of years (but see e.g. Zhou et al., 2008). A similar but 'reverse' mechanism is found in birds, which have a female (ZW) – male (ZZ) system that is conserved across almost all bird species (but see e.g. Pala et al. 2012). Considering these classical models, the main driving force behind the evolution of sex chromosomes is sexually antagonistic selection. When males and females have different evolutionary interests, i.e. if one behaviour or phenotypic trait is favoured in males but not females, allele frequencies will start to diverge between the sexes, creating what is termed 'sexual conflict' (Rice 1984). Sex chromosomes evolve from autosomes when one or multiple loci acquire the role of 'master sex-determining gene(s)'. Such loci can emerge from mutations in introns or regulatory elements, or gene duplication with neofunctionalization within the sex-regulatory cascade (Furman et al. 2020; Pan et al. 2021). Other variants can then hitchhike with the sex-determining gene if they are advantageous to one of the sexes and disadvantageous to the other (Van Doorn and Kirkpatrick 2007; Furman et al. 2020). This process suppresses recombination by selecting against recombinant genotypes, increases sequence divergence between sex-linked regions on the X and Y or Z and W chromosomes, and extends the sex-linked regions. Over time, this causes degeneration of the Y or W chromosome (= heteromorphy, through accumulation of mutations and gene loss). The sexual conflict is solved because females (XX) or males (ZZ) no longer

carry the male (Y) or female (W) beneficial alleles, respectively (Furman et al., 2020). Many variations of the well-known XX/XY and ZZ/ZW systems exist. There are cases in which the male or female carries one chromosome less than the other sex (XX/XO or ZO/ZZ, respectively) (Furman et al., 2020). On the other side of the spectrum, there are multiple sex chromosome systems, e.g. $X_1X_1X_2X_2/X_1X_2Y$, XX/XY_1Y_2 or ZW_1W_2/ZZ. Such systems can be formed through fusions of sex-chromosomes (most often the Y) with autosomes, centric fission of the X chromosome, translocation of an ancestral sex-determining locus to an autosome, or emergence of a new sex-determining locus on an autosome (White 1973; Van Doorn and Kirkpatrick 2007; Kitano and Peichel 2012). Sex can also be determined by external factors, such as temperature or social cues, which is often the case in fishes and reptiles (Korpelainen 1990; Ezaz et al. 2006). Next to clades with separate sexes, there are hermaphroditic species that experience a sex change at some point of their life (Mank et al. 2006).

Contrary to the stability in mammals in birds, sex chromosome systems in fishes, reptiles and amphibians are extremely variable. In these classes, switches of the location of the sex-linked region, as well as switches between XX/XY, ZZ/ZW, multiple sex chromosome systems and environmentally dependent sex-determination have occurred often and independently (Ezaz et al. 2006). Sex chromosome systems can differ between closely related fish species (Korpelainen 1990; Ezaz et al. 2006; Henning et al. 2008; Ross et al. 2009; Kitano and Peichel 2012; Cioffi et al. 2013; Myosho et al. 2015), and even between populations or individuals of the same species (Miura et al. 1998; Volff and Schartl 2002; Rosa et al. 2009; Reichwald et al. 2015; Triay et al. 2022). Mechanisms of sex chromosome switches include the emergence of a novel sex-determining mutation on an autosome, or translocation of a sex-determining locus from a sex chromosome to an autosome (non-homologous turnover). The same events can also occur within an existing sex chromosome (homologous turnover). Depending on how dominant the new sex locus is over the ancestral one, a multifactorial system may persist for some time before a switch (Furman et al. 2020) or even remain stable (Van Doorn and Kirkpatrick 2007). Some genes have a tendency to repeatedly and

independently emerge as new sex-determiners, but truly novel, unexpected candidates are also constantly found (Furman et al. 2020; El Taher et al. 2021; Pan et al. 2021).

The huge diversity of sex chromosomes and sex-determining mechanisms in fishes begs the question what role sex chromosome turnover plays in fish speciation. In heteromorphic sex chromosome systems, recombination in and around sex-linked sequences is suppressed in the heterogametic or hemizygous sex. Because of the hemizygosity (XY males or ZW females with degenerate Y or W chromosomes), recessive beneficial mutations are exposed to selection in hemizygous (XY or ZW) individuals, which would otherwise be masked by heterozygosity. In addition, effective population size is smaller and genetic drift higher in Y and W chromosomes than in autosomes. These regions are thus expected to diverge faster between differentiating populations in response to selection and are expected to disproportionally contribute to pre-zygotic and post-zygotic isolation (Payseur et al. 2018; Dufresnes and Crochet 2022). Reproductive barriers are needed for speciation to occur. If a population contains individuals with diverging sex-determining mechanisms, hybrids between such individuals may be infertile, unviable or have lower fitness, which can form a barrier on its own. In addition, other variants causing hybrid sterility or behavioural isolation (e.g. mate recognition) easily accumulate on sex chromosomes due to the strong recombination suppression in heteromorphic sex chromosome (Lindholm and Breden 2002; Qvarnström and Bailey 2009). However, heteromorphy is relatively rare in fishes (Devlin and Nagahama 2002). The role of sex chromosome turnover in fish speciation is supported by correlations between species diversity and rate of sex chromosome turnover (El Taher et al. 2021) and the accumulation of genes for hybrid sterility and behavioural isolation on sex chromosomes (Qvarnström and Bailey 2009). For instance, the formation of a new sex chromosome has allowed reproductive isolation to evolve in sympatry between Pacific and Japanese three-spined sticklebacks. In this incipient species, genes for male courtship display are linked to the neo-Y chromosome (Kitano and Peichel 2012). A similar case was recently found in nine-

spined sticklebacks, where a new sex-determining system (XY compared to ancestral ZW) evolved in some populations, but not in others, and has caused hybrid sterility between them (Natri et al. 2019).

Once the initial suppression of recombination in a sex-linked region is established, we can expect selection for other mechanisms causing recombination suppression. Mechanisms that are favoured in sex-linked regions include series of inversions, the insertion of transposable elements, and high levels of DNA methylation. Aside from adding more recombination suppression, the ‘normal’ level of selection against maintaining original gene order and insertion of additional transposable elements is relaxed, allowing the accumulation of these features on sex chromosomes (Furman et al. 2020). Evidence for interactions between sex chromosomes and inversions is ample. Fixation of inversions is favoured on homogametic (X or Z) sex chromosomes compared to autosomes, while inversion polymorphism is more often maintained on autosomes, due to the selection for additional recombination suppression and due to a heavier load of deleterious mutations on autosomes (Connallon et al. 2018; Furman et al. 2020). Sex chromosomes also have a smaller effective population size, which further facilitates the purging of deleterious recombinant genotypes (Qvarnström and Bailey 2009; Kirkpatrick 2010). In birds, inversions more frequently occur within the Z-chromosome compared to autosomes, and fixation of such inversions contributes disproportionately to differentiation between subspecies in hybrid zones (Hooper and Price 2017). A novel sex determination system in some nine-spined stickleback populations emerged inside of a large inversion on LG12, causing sterility of hybrids between two lineages (Natri et al. 2019). In its close relative, the three-spined stickleback, the XY sex-determination system is also contained by three stratified inversions but on a different chromosome (LG19, Peichel et al., 2020). If a new inversion captures a part of a sex-determining region and a sex-antagonistic gene, they become tightly linked and can transfer the primary sex-determination system to this new inverted region (Van Doorn and Kirkpatrick 2007). In humans, where the Y-chromosome is much older than in sticklebacks, a series of inversions kept adding loci

to existing non-recombining blocks, extending the sex-linked region (Lahn and Page 1999).

1.4 The African Great Lakes: a study system for fundamental and applied fish research

1.4.1 Geological history and diversification of fish lineages of the African Great Lakes

The African Great Lakes have been model systems for the biogeography, ecology and evolutionary biology of freshwater fishes for over a century (Boulenger 1898). They are also called the Rift Valley Lakes, as they were formed by tectonic activity starting 40 millions of years ago (MYA), although the largest of the lakes are only around 10 MY old (Tiercelin and Lezzar 2002). According to the definition of Salzburger et al. (2014), in which an African Great Lake must cover > 2,000 km^2 in area, there are nine lakes; the three largest ones (Victoria, Tanganyika, and Malawi) and six smaller ones (Turkana, Albert, Mweru, Rukwa, Kivu, and Edward) (Figure 2). Especially the three larger lakes harbour an immense diversity of fishes, ostracods, gastropods, and bivalves with an exceptionally high percentage of endemism (Salzburger et al. 2014), explained by the near-isolation of African freshwater habitats from other continents since the Cretaceous (Lavoué 2020). The adaptive radiations of cichlids are the mantlepiece of this diversity, most of which are unique within lakes and show an astonishing reservoir of morphological and behavioural adaptations (Salzburger et al. 2014). The colonization history of African Great Lake fish lineages is shaped by the turbulent tectonic and hydrological history of the continent. Temperatures and global sea levels rose and fell drastically during the last 100 MY, creating and breaking connections between African water bodies, the Atlantic and Pacific Ocean, and the water bodies of the Arabian plate. Thus, many fish lineages in the African Great Lakes may be traced back to marine origins, but many also crossed via land from the Orient (Lowe-Mcconnell 1993; Lavoué 2020).

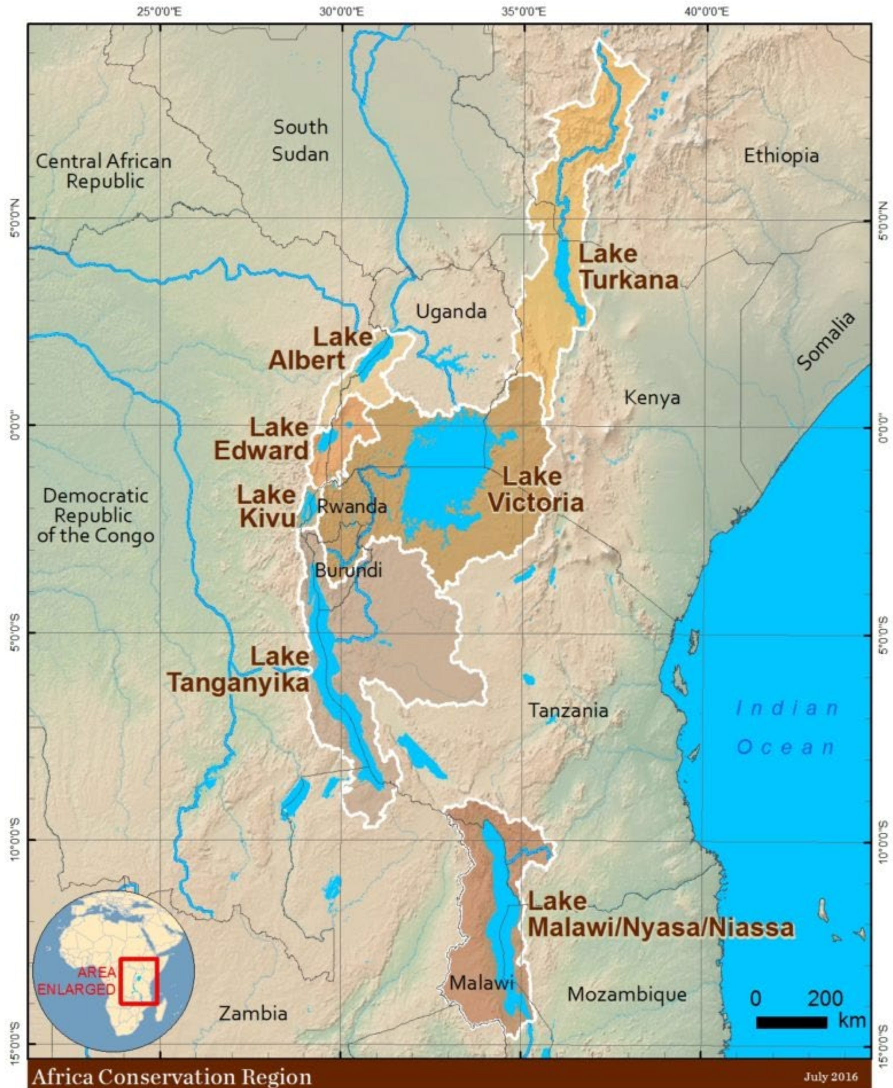

Figure 2. Map of the African Great Lakes region. From https://www.greatlakesofafrica.org/about-the-lakes/

1.4.2 Small pelagic fishes in the African Great Lakes

Most textbook examples of small pelagic fishes are marine (e.g. the species discussed in section 1.2 such as Atlantic herring, European sardine, anchovy, …). However, given the large area and depth of the African Great Lakes (Salzburger et al. 2014), their pelagic realms harbour productive fisheries of small pelagic fishes, which are essential

for food security in Central and East-Africa. Small pelagic fishes are highly nutritious (protein, micronutrients), usually processed (sun-dried) and consumed whole, but also used as fishmeal for aquaculture. The majority of these fisheries are small-scale or semi-industrial, little regulated with a high incidence of illegal fishing, and underreported and undercontrolled for quality with major losses to spoilage (Wangechi et al. 2015; Kolding et al. 2019; FAO 2023). The small pelagic fishes in the African Great Lakes are taxonomically diverse and belong to five families; Cyprinidae (carps), Cichlidae (cichlids), Dorosomatidae (gizzard shads and sardinellas), Alestidae (African tetras), and Danionidae (danionins) (Table 1). They are exploited at different levels (industrial, artisanal, or subsistence fishing) by a multitude of different fishing gears including trawls, gill nets, mosquito nets, and lift nets. Lift net fishing is highly efficient because non-motorized or motorized vessels set out at night and attract the plankton-feeding fishes with powerful lights. All commercially exploited small pelagics of the Great Lakes share the typical r-selected characteristics and behaviours presented in section 1.2.1, although there is some variation in feeding and spawning behaviour (Kolding et al. 2019).

Table 1. Overview of small pelagic fishes in the African Great Lakes. Local names and yield estimates are from Kolding et al. (2019) unless otherwise indicated. Habitat and distribution information are from FishBase (Froese and Pauly 2024).

Lake	Area (km^2) / maximum depth (m)	Family	Species	Local name	Potential (P) or realized (R) yield ($t \times y^{-1}$)	Habitat	Distribution
Victoria	59,947 / 81	Cyprinidae	*Rastrineobola argentea*	Dagaa, omena, mukene	509,120 (R)	Pelagic	Lakes Victoria, Kyoga, Nabugabo, Victoria Nile
Tanganyika	32,900 / 1435	Dorosomatidae	*Stolothrissa tanganicae*	Dagaa, kapenta	111,860 (R)	Pelagic	Lake Tanganyika, Lukuga river
			Limnothrissa miodon	Dagaa, kapenta		Littoral + pelagic	Lakes Tanganyika, Kivu, Itezhi-Tezhi, Kariba, Cahora Bassa
Malawi	29,600 / 704	Cyprinidae	*Engraulicypris sardella*	Usipa	54,801 (R)	Demersal (littoral) + pelagic	Lake Malawi, upper Shire River
		Cichlidae	Mixed Haplochromini	Utaka	–	Littoral + pelagic	
Turkana	6,405 / 109	Alestidae	*Alestes baremoze*	Lokabela	–	Benthopelagic, potamodromous	Senegal, Gambia, coastal basins of Côte d'Ivoire, Volta, Niger/Benué, Chad basin, Nile Omo, lakes Albert and Turkana
		Alestidae	*Brycinus minutus*	Lokabela	300.000 (P)	Pelagic	Lake Turkana
		Alestidae	*Brycinus ferox*	?		Pelagic	Lake Turkana

Albert	5,590 / 51	Cyprinidae	*Engraulicypris bredoi*	Muzizi	128.000 (R)[1]	Benthopelagic	Lake Albert
		Alestidae	*Brycinus nurse*	Ragoogi	32.000 (R)[1]	Pelagic, potamodromous	West Africa, Nile river up to Lake Albert
Mweru	5,120 / 27	Dorosomatidae	*Microthrissa moeruensis*	Chisense	67,000 – 95,000 (P)[2] 50,176 (R)	Pelagic	Lake Mweru
		Cyprinidae	*Engraulicypris moeruensis*	Chisense	–	Benthopelagic	Mweru-Luapula, Bengweulu-Chambeshi
		Danionidae	*Chelaethiops congicus*	?	–	Benthopelagic	Congo river basin, Lake Tanganyika basin
Kivu	2,700 / 480	Dorosomatidae	*Limnothrissa miodon*	Kapenta	21,400	Littoral + pelagic	Lakes Tanganyika, Kivu, Itezhi-Tezhi, Kariba, Cahora Bassa
Edward	2,325 / 112	Cichlidae	Mixed Haplochromini	?	–		
		Cyprinidae	*Engraulicypris* spp.	?	–		
		Alestidae	*Brycinus nurse*	?	–	Pelagic, potamodromous	West Africa, Nile river up to Lake Albert
Rukwa	2,000 / 15	Cyprinidae	*Engraulicypris spinifer*	?	–	Benthopelagic	Malagarazi river, Ruaha river, Lake Rukwa

References: [1]Nakiyende et al. (2018), [2]van Zwieten et al. (1996)

Some of the fisheries on small pelagic fishes in the African Great Lakes are well-developed, like the fishery of the cyprinid *Rastrineobola argentea* (local names: 'dagaa', 'omena' or 'mukene') in Lake Victoria (Lowe-Mcconnell 1993; Kolding et al. 2019), the dorosomatids *Stolothrissa tanganicae* and *Limnothrissa miodon* ('dagaa' or 'kapenta') in Lake Tanganyika and Kivu (Coulter 1991; Mölsä et al. 1999; Kolding et al. 2019), and the cyprinid *Engraulicypris bredoi* ('muzizi') and the alestid *Brycinus nurse* ('ragoogi') in Lake Albert (Nakiyende et al. 2018). Others, such as the fishery of *Engraulicypris sardella* ('usipa') in Lake Malawi and that of *Microthrissa moeruensis* ('chisense') are currently still underexploited (Kolding et al. 2019; Simfukwe et al. 2022). Still others are only locally fished and poorly characterized, for example *Engraulicypris moeruensis* (also part of 'chisense') and *Chelaethiops congicus* in Lake Mweru (Kolding et al. 2019), the haplochromide cichlids, *Engraulicypris* spp. and *B. nurse* fisheries in Lake Edward (Marshall 1984; Kolding et al. 2019; Nakiyende et al. 2018), the *Mesobola spinifer* subsistence fishery in Lake Rukwa (Kolding et al. 2019), and the *Alestes* spp. and *Brycinus* spp. fisheries in Lake Turkana (Kolding et al. 2019; Getahun et al. 2020) (Table 1). In several of the Great Lakes (Victoria, Malawi, Mweru, and Albert), small pelagic fishes emerged as the dominant fishing targets after the decline of larger species (Marshall 1984; Wangechi et al. 2015; Wangechi 2016; Nakiyende et al. 2018; Kolding et al. 2019).

Despite their commercial importance and in stark contrast with the multitude of studies on small pelagic fishes in the Global North (see sections 1.2.2 and 1.2.3), genetic and especially genomic studies on the small pelagic fishes in the African Great Lakes are still scarce. Nevertheless, some pioneering work has been done. Anseeuw (2011, 2014) found evidence for low but significant differentiation between populations of some of the pelagic haplochromine cichlid species in Lake Malawi. In Lake Victoria, Wangechi et al. (2015) found genetic evidence for recent population growth of *R. argentea.* The authors linked this to the decline of the native haplochromine cichlid community, which was competing with *R. argentea* for food. The combination of predation by the Nile perch, fishing pressure, and increasing extent of hypoxia has

induced rapid evolution in *R. argentea.* It has expanded its niche from littoral and pelagic to more hypoxic benthic waters and adopted a different diet, resulting in increased head size and gill surface area, and decreased gill raker size (Wangechi et al. 2015; Sharpe and Chapman 2018). In parallel, fisheries-induced evolution has reduced body size, size at maturity, and increased reproductive effort in *R. argentea* (Sharpe et al. 2012; Wangechi et al. 2015). Lake-wide genetic population structure is still unkown, but there is some indication of local fine-scale structure (Wangechi 2016). More evidence for population structuring comes from morphological studies. Lake Mweru's *Microthrissa moeruensis* ('chisense') has two morphotypes, of which one is fully zooplanktivorous and the other partly insectivorous, resembling the ecological separation of *S. tanganicae* and *L. miodon* in Lake Tanganyika (see section 1.4.3). The two morphotypes of *M. moeruensis* could represent a case of incipient speciation, but are not geographically separated and are not sufficiently genetically differentiated to separate as two species (Strømme 1999). *Alestes baremoze* from Lake Turkana has also been subdivided into three morphological subspecies (Blache 1964; Paugy 1986), but based on ecological similarities*,* they probably represent subpopulations of the same species as well (Froese and Pauly 2024).

1.4.3 The Lake Tanganyika sardines

The main study system of this thesis, Lake Tanganyika, harbours the most well-known and well-developed pelagic fishery among the African Great Lakes. It is dominated by two species of clupeids (*Stolothrissa tanganicae* and *Limnothrissa miodon*), locally called 'dagaa' or 'kapenta', which together yield an estimated catch of 111,860 $t \times y^{-1}$, feeding millions of people in the riparian countries of the lake (Kolding et al. 2019) (Figure 3). Despite their apparent similarity, the sardines possess species-specific characteristics that may cause them to react differently to climate change and fishing pressure. For instance, *S. tanganicae* lives a fully pelagic life, while *L. miodon* spawns in the littoral zone and only moves to the pelagic zone after reaching a certain length (Mannini et al., 1996; Mulimbwa et al., 2022). Individuals of *L. miodon* also grow larger

and show more flexible feeding behaviour than those of the completely planktivorous *S. tanganicae* (Mulimbwa & Shirakihara, 1994; Mulimbwa et al., 2014).

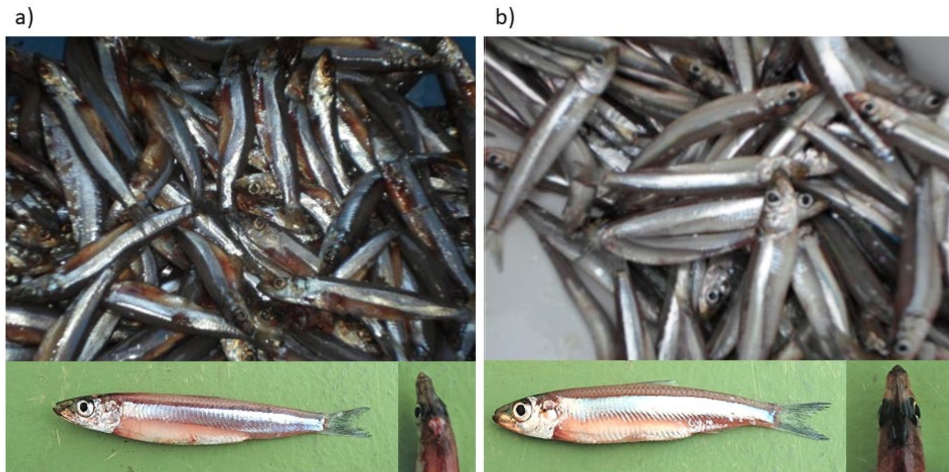

Figure 3. Photographs of the Lake Tanganyika sardines a) *Stolothrissa tanganicae* and b) *Limnothrissa miodon*, showing their morphological differences. Adapted from Abdulkarim et al. (2014) and Pekka Kotilainen.

Given the large size of Lake Tanganyika and the fact that it is surrounded by four countries (Democratic Republic of the Congo, Burundi, Tanzania, and Zambia), an important management question is whether the sardines form a single panmictic stock or separate stocks, i.e. according to the lake's sub-basins or some other geographical or ecological separation. Hence, the population structure of the Tanganyika sardines has been subject to a number of studies. Early studies using restriction fragment length polymorphisms (RFLP) and otolith markers did not find any consistent structure (Kuusipalo 1999; Sako et al. 2005). More recent studies unveiled a weak pattern of isolation-by-distance (IBD), which was more prominent in *L. miodon*, possibly reflecting its littoral affinity. In addition, a large non-geographically segregating inversion was found in *L. miodon* (De Keyzer et al. 2019; Junker et al. 2020). Inversion frequencies also differed between Lake Tanganyika and Kivu (Junker et al., 2020). Populations of the latid and cichlid predators of the sardines are also panmictic (Koblmüller et al. 2019; Rick et al. 2022).

As the niche of pelagic planktivores was empty in several lakes in and around the Rift Valley, several thousand juvenile clupeids were introduced into three other lakes for fishery purposes: the natural Lake Kivu in DR Congo in 1958 and 1960 (Collart 1960; Spliethoff et al. 1983), the man-made Lake Kariba in Zambia and Zimbabwe between 1967 and 1968 (Bell-Cross and Bell-Cross 1971; Junor and Begg 1971), and the man-made Itezhi-Tezhi in Zambia in 1992 (Mubamba 1992). Only *L. miodon* established and even spread via the Zambezi river into the Cahora Bassa reservoir in Mozambique downstream of Lake Kariba (Bernacsek and Lopes 1984). In all cases, it now supports productive fisheries (Bernacsek and Lopes 1984; Marshall 1988, 1995; Kolding et al. 2019; Tessier et al. 2020). The major documented ecological effects of these introductions are shifts in the local zooplankton community through predation, the unintentional introduction of the Tanganyika killifish *Lamprichthys tanganicanus* into Lake Kivu, and the co-introduction of the gill-infecting monogenean flatworm *Kapentagyrus limnotrissae* into Lake Kariba (de Iongh et al. 1983; Dumont 1986; Marshall 1991; Muderhwa and Matabaro 2010; Kmentová et al. 2019). The ecological success and limited side effects can be attributed to the relatively species-poor fish fauna of the lakes in which *L. miodon* was introduced (Ogutu-Ohwayo and Hecky 1991; Lévêque 1996). The reservoirs are relatively species-poor because of their recent damming and depletion of local fauna, while Lake Kivu is relatively species-poor because of its recent volcanic formation, as well as the waterfalls on the Ruzizi river that prevent colonization from Lake Tanganyika (de Iongh et al. 1983). Introduction of Tanganyika sardines into Lake Malawi was also proposed (Turner 1982). This was advised against because the sardines would disrupt the complex trophic networks of the local cichlids, directly compete with other pelagic zooplanktivore species, introduce pathogens, and require an expensive switch of fishing gear (Eccles 1985; Ogutu-Ohwayo and Hecky 1991).

1.4.4 A brief history of fisheries management in Lake Tanganyika

The four countries surrounding Lake Tanganyika have long recognized the importance of harmonizing the management of its fisheries. In 1992, the Lake Tanganyika Research

(LTR) program was established between the national fisheries authorities and research institutes from the four states (Département des Eaux, Pêches et Pisciculture, Burundi; Service National de Développement de la Pêche and Centre de Recherche en Hydrobiologie, Democratic Republic of the Congo; Department of Fisheries and the Tanzanian Fisheries Research Institute, Tanzania; and Department of Fisheries, Zambia), executed by the Food and Agriculture Organization of the United Nations (FAO) and scientifically coordinated by the University of Kuopio, Finland. At this time, the fisheries of Lake Tanganyika were essentially 'open access' within each country's national borders, which was deemed entirely unsustainable. The LTR formulated a fisheries management framework in 1999 (Reynolds 1999). In this plan the LTR placed much emphasis on a holistic approach to fisheries management, in particular towards inclusion of social aspects and co-management involving local communities. They also called for efforts to implement the proposed measures in a unified manner across all four countries. Management guidelines were based on the FAO's reference framework, the 'Code of Conduct for Responsible Fisheries' (FAO 1995), following their guiding principles; sustainability, best scientific evidence, participation and cooperation, objectivity and transparency, timeliness, and flexibility. Further complementary studies were launched under the Lake Tanganyika Biodiversity Project (LTBP), which focused on pollution control, conservation, and the maintenance of biodiversity (UNDP 1994).

Together, the research efforts collated in a legal framework drafted by the newly established Convention on the Sustainable Management of Lake Tanganyika (further referred to as 'the convention') (FAO 2003). In this treaty it was agreed that the lake's resources, including fisheries, should be managed in a way that is precautionary, preventative, fair and equitable, adaptive, and does not negatively affect the other countries ("prevention and minimization of transboundary impacts"). The convention established the Lake Tanganyika Authority (LTA) as an executive and coordinating unit, which started operating under the Lake Tanganyika Regional Integrated Management Programme (LTRIMP) in 2008. The convention's strategic action plan highlighted the

common problem for the pelagic fisheries of all four countries: excessive and/or uncontrolled offshore fishing. The LTA continues to meet annually to update on scientific and applied progress, the issues that arise, and set priorities for the coming year. Decisions on the sustainable management of the lake are discussed by representatives of all four countries together (LTA 2009). Other short-term monitoring projects include CLIMLAKE/CLIMFISH by BELSPO (2002-2006) and CHOLTIC (2011-2014). Regional agencies also implement their own monitoring programmes, but this information is not integrated at the lake level. Despite all efforts, a standardized, continuous long-term monitoring program that involves all four countries around Lake Tanganyika is still lacking, which can cause stakeholder disagreements about the state of natural resources and contradictory actions (Plisnier et al. 2018).

2 Overarching goal & aims

The Lake Tanganyika sardines offer an opportunity to simultaneously study fundamental principles of evolution of pelagic fishes, and aid the development of sustainable, science-based fisheries practices in this study system and beyond. The overarching goal of this thesis was thus to gain better understanding of the ecological and genomic factors that have played a role in the speciation and contemporary evolution of these fishes, and to put these insights in a fisheries context.

The specific aims of this thesis were to:

1. Develop genomic resources for the Lake Tanganyika sardines (**Paper I & II**)
2. Describe the the phylogenomic/geological context for their divergence (**Paper I**)
3. Within this context, gain better understanding of the genomic and ecological factors that played a role in their divergence and contemporary evolution (**Paper I, II & III**)
4. Based on our genomic data as well as insights from surveys with fish and fisheries experts, formulate recommendations for fisheries management of the Lake Tanganyika sardines (**Paper II, III & IV**)

3 Main findings and discussion

3.1 New genomic resources for the Tanganyika sardines (Paper I & II)

In **Paper I** and **Paper II** of this thesis, we produced large volumes of different types of genomic data (DNA, RNA, short, long, and long-linkage reads) and used them to develop new genomic resources for the two clupeid species of Lake Tanganyika; *Stolothrissa tanganicae* and *Limnothrissa miodon.* First, we assembled and annotated their mitochondrial genomes from shotgun NGS reads using a reference-based approach. The new mitogenomes were comparable in gene content, length and structure to other clupeiforms and teleosts (Satoh et al. 2016). Second, we assembled high-quality, near chromosome-level nuclear genomes based on three levels of long-range linkage information (10X, Chicago, and Hi-C) and annotated them using newly generated long-read RNA data. Chromosome-level assemblies are the gold standard of genome assembly and are becoming more attainable thanks to the combination of short-read, long-read, and long-range linkage technologies. The high quality of the new genomes is demonstrated by comparison to the genomes of other commercially important fish species. The N50 values, a measure of contiguity, of the *S. tanganicae* and *L. miodon* assemblies are 53.5 and 35.5 Mb, respectively. Those of other commercially important species range between 0.4 Mb (rainbow trout) and 25.8 Mb (Asian seabass) (Lu and Luo, 2020). Scores of genome completeness (BUSCO) for *S. tanganicae* and *L. miodon* were 93.4 % and 92.4 %, respectively, which is somewhat lower than other exploited clupeiforms, e.g. *Clupea harengus* (97.5 %, accession: GCF_900700415.2), *Alosa sapidissima* (98.5 %, accession: GCF_018492685.1), and *Denticeps clupeoides* (98.2 %, accession: GCF_900700375.1). A recent chromosome-level assembly of *Setipinna enuifilis* which also used Hi-C data had an N50 of 32.42 Mb and a BUSCO score of 89.6% (Liu et al., 2023). Genome size is very similar between these and other clupeiforms and the Tanganyika sardines and lies around 0.8 Gb (Liu et al., 2022; Ma et al., 2023).

The new (mito)genomic resources generated in this thesis allowed analyses that were previously impossible. We used the new mitogenomes to construct the first whole-mitogenome phylogeny of Clupeiformes that included the Tanganyika sardines (**Paper I**). Previous phylogenies have only included one to three mitochondrial genes for *S. tanganicae* and *L. miodon.* Due to a higher total number of characters and accounting for diverse evolutionary characteristics of different genes, whole mitogenomes may be more appropriate than single-gene datasets when attempting to resolve phylogenies and date the divergence of recently diverged taxa (Cummings et al. 1995; Zardoya and Meyer 1996; Duchêne et al. 2011). Using the whole-mitogenome approach we were able to update their divergence time and outline a biogeographic context for their speciation (see section 3.2), as well as highlight phylogenetic inconsistencies in the pellonuline lineage.

The assembly and annotation of the nuclear genomes of the Tanganyika sardines opened up even more new avenues. First, reference genomes facilitate the comparison of individuals and populations at a genome-wide scale, thereby greatly increasing the number of markers compared to classic genetics. This in turn improves the resolution of population structure analyses and the accuracy with which demographic parameters can be estimated (Primmer 2009). Second, comparing allele frequencies of highly differentiated single nucleotide polymorphisms (SNPs) to a whole-genome background allows the identification of adaptive variation, which can aid in identifying biologically relevant management units (but see next paragraph) (Hauser and Carvalho 2008). Third, mapping population-level variation to scaffolds, pseudo-chromosomes or linkage groups removes anonymity from the markers. It links them to specific locations in the genome, to each other, and, if an annotation is available, to genes and functions (Allendorf et al. 2010). Fourth, annotated whole genomes enable comparative genomic analyses, such as establishment of regions with conserved gene order (synteny) and identification of rearrangements (Sarropoulou and Fernandes 2011). We leveraged the power of the new genomic resources and mapped lake-wide reduced-representation data (RAD-seq) to the new genome assemblies, which allowed us to identify tightly

linked regions of high differentiation between and within species (**Paper II**). First, we delineated the sex chromosomes of both species, and confirmed previous findings that the Tanganyika sardines have a different sex chromosome organization (Junker et al. 2020). The new annotations enabled characterization of the genes in these regions, and showed that the highly sex-differentiated gene sets also largely differed between the species. Second, we delineated a large inversion (9.7 Mb) in *L. miodon,* adding to the results of Junker et al. (2020) by showing that there is variation in inversion group frequencies along the North-South axis of Lake Tanganyika. Linking the highly differentiated loci to genes, we suggest a role of the inversion in local adaptation to limnological conditions through vision and gamete properties. Third, the annotation allowed us to carry out gene-level synteny analyses and establish conserved and rearranged regions between the genomes of the Tanganyika sardines and other clupeids (**Paper II**). Finally, the genome of *L. miodon* was used to map RAD-seq data from four other lakes in which *L. miodon* was introduced since the 1960s, enabling us to estimate effective population size, genetic differentiation, and signatures of selection between the lakes (**Paper III**). Using diagnostic loci defined in **Paper II,** we also showed that frequencies of the inversion karyotypes differ not only between the sub-basins of Lake Tanganyika, but also between the lakes Tanganyika, Kivu, Itezhi-Tezhi, Kariba and Cahora Bassa. This indicates a possible role of the inversion in between-lake adaptation. The causal role of the inversion in adaptation needs to be verified, for instance by examining sequence similarity in inversion-linked genes between locations or by correlating the karyotypes and genotypes to environmental variables rather than just distance and isolation (Baltazar-Soares et al. 2018).

One of the greatest advantages of annotated reference genomes is the possibility to characterize both neutral and adaptive genetic diversity (see introduction section 1.1.2). Arguably, genetic variants that are under selection because of their link to a functional trait, such as alternative life histories (Petrou et al. 2021), migration behaviour (Hess et al. 2013) or ecological preferences (Berg et al. 2015; Barth et al. 2017), are more relevant for the definition of management units than those that only

reflect demographic history (Baltazar-Soares et al. 2018). Identifying such variants in the first place can be more time-consuming and expensive, especially if the functional links remain to be proven (e.g. with experiments, QTL, candidate gene approaches, …) (Gagnaire and Gaggiotti 2016; Bernatchez et al. 2017). However, once established, they can easily become part of routine monitoring programmes using the same sampling and laboratory techniques (Baltazar-Soares et al. 2018). Even without these additional steps, if strongly differentiated loci are interpreted as a proxy for adaptive genetic diversity, it may be quite feasible to identify regions under putative selection. These can then be mapped to reference genomes and annotations of the target species or related species in publicly available databases. Proving more than correlation requires a more elaborate approach, e.g. combining biophysical models, experiments, and genome scans (Baltazar-Soares et al. 2018). In small pelagic fishes, the proxy approach is aided by the low overall levels of differentiation (Hauser and Carvalho 2008), where loci or regions ('islands of selection') sticking out from the low differentiation background are good indicators for selection (Bernatchez et al. 2017). Nevertheless, genome scans have their inherent pitfalls, and it is often not straightforward to distinguish between neutral loci and loci under weak or balancing selection, or loci with small effect (Gagnaire and Gaggiotti 2016; Bernatchez et al. 2017). In addition, reduced representation approaches like the one we used (**Paper II & III**) are prone to missing many loci under selection, especially in species without a reference genome (Lowry et al. 2017).

High-quality genomes can answer numerous other micro- and macroevolutionary questions. Particularly interesting future avenues are the analysis of repetitive elements and their potential role in adaptation, the role of transposable elements in the chromosomal rearrangements in the clupeiform lineage, gene family expansions and contractions in comparison to other pellonulines or clupeids (e.g. in relation to marine-freshwater transitions, their diversification across Africa and riverine-lacustrine transitions), and coalescence approaches within species over evolutionary timescales linked to changes in climate and anthropogenic pressures.

3.2 Biogeographic context: diversification across Africa and intralacustrine speciation (Paper I)

Clupeiformes is a diverse, rapidly evolving and adaptable clade that has colonized the entire globe and various aquatic habitats. Transitions from marine to freshwater and between marine and euryhaline lifestyles have occurred repeatedly and independently across the entire lineage (Lavoué et al. 2013; Bloom and Egan 2018). Switches between resident freshwater and anadromous or catadromous lifestyles have also been documented (Bloom and Lovejoy 2014; Bloom and Egan 2018), as well as a diversity of diets (zooplanktivore, herbivore, insectivore, molluscivore, and crustacivore) even between closely related species (Egan et al. 2018). Switches from riverine to lacustrine conditions are known as well, but have not yet been systematically analysed. For example, the tremendous freshwater fish diversity of South-America has been shaped by eustatic sea level changes (50-10 MYA) and geological uplifts, intermittently connecting and isolating low-elevation plains and basins and upland rivers (Cassemiro et al. 2023). Similarly, the tribe Pellonulini, endemic to Africa, is thought to derive from marine ancestors and has spread throughout Central-Africa via the Congo river basin around 50 MYA (Wilson et al. 2008; Lavoué 2020). Pellonulini is a primarily freshwater tribe, with some euryhaline members and a euryhaline (catadromous) sister species, *Ethmalosa fimbriata* (Whitehead 1985; Lévêque et al. 1990). Their diversification was most likely driven by repeated isolation and reconnection of the Congo River and its tributaries and associated water bodies, due to geological events and during periods of rising and falling sea levels (Haq et al. 1987; Van Sickel et al. 2004).

The Tanganyika sardines are among the most recently formed and genetically similar species in the pellonuline phylogeny (**Paper I**). Our mitogenomic phylogeny estimated the split between *S. tanganicae* and *L. miodon* at 3.64 [95% CI: 0.99–6.29] MYA, a relatively young estimate compared to other studies (Wilson et al. 2008; Lavoué et al. 2013; Bloom and Lovejoy 2014; Egan et al. 2018). The divergence time from their closest relative, *Potamothrissa obtusirostris,* was estimated at 10.92 MYA [95% CI:

6.37–15.48]. Given these estimates, we suggest that during the formation of Lake Tanganyika from the Malagarasi-Congo River (around 9-12 MYA), a riverine generalist ancestor entered the proto-lake Tanganyika and that speciation of *S. tanganicae* and *L. miodon* was triggered by the fusion of the sub-basins and the onset of clearwater conditions (around 5-6 MYA) (Cohen et al. 1993). An alternative explanation, which is not mutually exclusive given the confidence interval, is that (repeated and/or partial) allopatry aided the speciation process during periods of drought and low lake levels (Cane and Molnar 2001).

The biogeographic history we outline for the Tanganyika sardines (**Paper I**) is comparable to other fishes, both on the African scale (pre-Tanganyika) and on the lake scale (post-Tanganyika). On the African scale, other lineages, such as Kneriidae and Phractolaemidae, show similar present-day distributions with marine origins (Lavoué 2020), although the marine-to-freshwater transition occurred tens of millions of years earlier in these two families (Lavoué et al. 2012). The diversification in Africa of some other taxa with a marine origin is estimated much younger than that of the sardines. For instance, *Lates*, known as the lates perches, also has representatives in Lake Tanganyika which overlap in their distribution with Pellonulini, with additional occurrences in the Orient (Lavoué 2020). However, their diversification in Africa is estimated at around 10 MYA (Koblmüller et al. 2021). On the lake scale, diversification events of fish lineages have been linked to the initial formation of Lake Tanganyika, the later formation of the northern (7-8 MYA) or southern (2-4 MYA) sub-basin, and the fusion of the then-present sub-basins and the onset of clearwater conditions (5-6 MYA). Around 3-4 MYA Africa experienced arid conditions (Cane and Molnar 2001), which may have impacted lake levels, (partially) separated the sub-basins and created opportunities for allopatric divergence. In addition, drill cores from Lake Malawi suggest periodically extremely dry conditions in the last million years, at times causing Lake Malawi to nearly disappear (Scholz et al. 2007; Ivory et al. 2016). Data for Lake Tanganyika are not available, but given its similar climate and latitude as Lake Malawi, water levels have likely fluctuated in parallel. The mostly littoral mastacembelids or

spiny eels radiated around 7-8 MYA, shortly after the lake's formation but before the establishment of fully lacustrine conditions (Brown et al. 2010). The speciation of some clades was dated around the time of the formation of the southern sub-basin (2-4 MYA) and/or low lake levels (Cane and Molnar 2001). The initial radiation of catfishes of *Synodontis,* a clade containing riverine, littoral, and benthic species happened 4-7.3 MYA in Lake Tanganyika at the onset of clearwater conditions. The two resulting endemic species flocks radiated later, around 2.5-3 MYA during a period of low lake levels (Day and Wilkinson 2006; Day et al. 2009). The four species of *Lates* that are endemic to Lake Tanganyika diversified around 2 MYA as well (Koblmüller et al. 2021). Rick et al. (2022) found weak and patchy population contemporary structure in three out of the four species of *Lates,* and suggested that it is linked to recent environmental fluctuations or spawning site fidelity.

3.3 Genomic and ecological factors underlying the divergence and contemporary evolution of the Tanganyika sardines (Paper I, II & III)

3.3.1 Factors contributing to between-species divergence: ecological opportunity, partial allopatry, and sex chromosome turnover

In freshwater fishes, geographic isolation is a major driver of speciation (Seehausen and Wagner 2014). Ecological and reproductive incompatibilities accumulate with or without divergent selection (Rundle and Nosil 2005). In contrast, the divergence time of the Tanganyika sardines suggests that they diverged in sympatry after the onset of clearwater conditions (**Paper I**). Heterogenous environments with plenty of ecological opportunities can trigger speciation without the need for physical isolation (Seehausen and Wagner 2014). The clearwater conditions presumably provided a new niche to be occupied and presented an opportunity for ecological speciation into a species with a more littoral affinity (*L. miodon*) and a fully pelagic one (*S. tanganicae*). As discussed in section 3.2, we cannot exclude that partial or full allopatry during arid periods have contributed to divergence between ancestral populations, which persisted upon

secondary contact. Differentiation between species was genome-wide, in contrast to the within-species differentiation in *L. miodon,* where it was concentrated in 'islands of divergence' (see section 3.3.2) (**Paper II**).

Another important factor in the evolution of the Tanganyika sardines are large-scale chromosomal rearrangements. While chromosome-scale synteny was observed between the Tanganyika sardines and other clupeids (*C. harengus* and *D. clupeoides*), we also found many small rearrangements. Fish genomes are remarkably diverse in size (Venkatesh 2003) and chromosome number (Mank and Avise 2006; Nirchio et al. 2014). Whole genome duplication events in teleosts complicate syntenic relationships, as they are followed by increased rate of rearrangements and gene family expansions (Venkatesh 2003; Sarropoulou and Fernandes 2011). Nevertheless, there is a high degree of synteny between the genomes of distantly related teleosts, even those with different chromosome numbers, with the more rearrangements the more phylogenetically distant they are (Sarropoulou and Fernandes 2011). A recent comparative analysis of three genomes of clupeiforms showed this pattern (higher collinearity between *Setipinna tenuifilis* and *C. harengus,* and lower collinearity between *S. tenuifilis* and *D. clupeoides*) (Liu et al., 2023). Liu et al. (2023) also indicated that chromosomal fusions and fissions are an important driver of chromosomal variation in the clupeiform lineage. Likewise, we observed an almost 1:1 relationship between *S. tanganicae* and *L. miodon* pseudo-chromosomes, with mostly small rearrangements, except between the sex chromosome of *L. miodon* and several autosomes of *S. tanganicae*. In addition, we found evidence for what is either chromosomal fusions or overscaffolding, see **Paper II** for a detailed discussion. Large syntenic blocks were also apparent in comparison to their more distant relatives *C. harengus* and *D. clupeoides*, but overall these were more fragmented and rearranged. Such rearrangements, generating variation in chromosome number, shape, and size, have been implied as drivers of speciation in several fish taxa, such as Elopiformes, Grammatidae, and *Coregonus* whitefishes (Molina et al. 2012; Symonová et al. 2013; de Sousa et al. 2021).

Another remarkable feature of the chromosomal evolution of the Tanganyika sardines is rapid sex chromosome turnover. Both species show a different sex chromosome organization compared to each other and to *C. harengus,* demonstrating that at least two switches have taken place in this lineage in the last 50 MY. Different sex chromosomes can form a post-zygotic barrier to hybridization between incipient species, but do not completely prevent it (Dufresnes et al. 2020). Indications for a role of sex chromosome turnover in speciation has been found in other fish lineages, such as cichlids and sticklebacks (Kitano and Peichel 2012; El Taher et al. 2021), but direct evidence remains scarce. Whether the different sex chromosomes have caused or aided reproductive isolation between Tanganyika sardines or their close relatives is still unclear, as we found little evidence for elevated between-species differentiation in the sex-linked regions (**Paper II**). This should not be confused with the different sets of genes implicated in between-sex differentiation in the two species.

3.3.2 Factors contributing to within-species divergence: migration, inversion-facilitated selection and (historical) allopatry

The evolution of every population is shaped by the forces of selection, gene flow, and genetic drift. The balance between these forces, and thus the evolutionary outcomes, can be altered by various ecological and genomic processes, such as physical isolation and reconnection, structural variation which alters the recombination landscape, or anthropogenic influences. As small pelagic fish species, the Tanganyika sardines are arguably mostly structured by high amounts of gene flow. Both species seem to migrate across large distances (Matthes 1967; Mölsä et al. 1999), presumably exchanging migrants across the entire lake, with a homogenizing effect on lake-wide genetic diversity. Several researchers have addressed the question whether there is population structure in either or both of the species, and whether one species is more spatially differentiated than the other. Genetic differentiation within species suggests that neither species is strongly differentiated, but the population of *S. tanganicae* in Lake Tanganyika is even less structured than that of *L. miodon* (**Paper II**). We suggest two non-mutually exclusive reasons. First, higher mobility and a more strictly pelagic

affinity in *S. tanganicae, versus* a more littoral affinity with regards to breeding (Mulimbwa et al., 2022), and potentially homing behaviour (breeding site fidelity) in *L. miodon* may restrict North-South gene flow in *L. miodon* compared to *S. tanganicae* (Kmentová et al., 2024). Second, the inversion in *L. miodon* may contribute to local adaptation to limnological conditions via opsin and gamete property pathways by suppressing recombination in the face of strong gene flow (**Paper II**). As a result, locally favourable alleles are protected from maladaptive alleles introduced from the other basins (Kirkpatrick and Barton 2006). The absence of the inversion in *S. tanganicae* suggests that it arose by chance in *L. miodon* but not *S. tanganicae*, or that it arose in an ancestor but one karyotype became fixed in *S. tanganicae.* The inversion might have aided *L. miodon* to adapt to local environmental conditions even if North-South migration was equally high in both species. The temporary disruption of gene flow may also have played an important role in the emergence and retention of the inversion polymorphism in *L. miodon.* As explained in section 3.2, periods of drought have caused lake-level fluctuations in Lake Tanganyika that temporarily separated the sub-basins and may allowed allopatric divergence (Tiercelin and Mondeguer, 1991). This may have driven the emergence and retention of the inversion polymorphism in *L. miodon*.

In the case of *L. miodon* that was introduced into other lakes (Kivu, Kariba, Itezhi-Tezhi, Cahora Bassa), gene flow was again, and this time permanently, disrupted. The non-native populations have been evolving in allopatry for decades, except those between Kariba and Cahora Bassa, where migration downstream the Zambezi river is still possible. Even though many generations have passed since the introductions/invasions, divergence between inversion groups was still larger than divergence between lakes (**Paper III**). Genetic differentiation between lakes was concentrated in sex-linked and inversion-linked regions, showing the powerful recombination suppression by inversions and on sex chromosomes (Hooper and Price 2017; Wellenreuther and Bernatchez 2018). Even if the different inversion karyotypes were not necessarily selected for, we may still see the pattern of recombination suppression (Rafajlović et al. 2021). Inversion frequencies also differed between lakes and partly corresponded

to the frequencies in the founding populations (‘northern Tanganyika’ karyotype frequencies in Lake Kivu, ‘southern Tanganyika’ karyotype frequencies in the Cahora Bassa reservoir). However, the pattern is not perfect, as the ‘southern’ karyotype is not the most frequent in Lake Itezhi-Tezhi and Kariba. The contemporary frequencies are most likely a combination of drift during and after stocking, and possibly selection for inversion-linked phenotypes. Chicoa basin in Cahora Bassa was the only location where divergence from other populations was stronger outside of the inversion, even within the same lake. This population is surprisingly distinct, and a combination of morphological, parasitological, and genomic studies including individuals along the Zambezi river connecting Kariba and Cahora Bassa could provide more insight into its origin (Rius et al. 2015).

3.4 Recommendations for fisheries management of the Tanganyika sardines and elsewhere based on two genomic studies and one social study (Paper II, II & IV)

How are the principles outlined by the LTA linked to the findings in this thesis? The genetic indicators with the highest management potential (**Paper II & III**) are 1) measures of genetic diversity and effective population size and 2) genetic population structure. In this section, we address each of them and formulate a recommendation based on the precautionary principle postulated in section 1.4.1 and in the fisheries management guidelines of the LTA. Using the findings of our survey of more than 100 African and non-African fish and fisheries experts (**Paper IV**), we also discuss the potential applicability of these recommendations in the socio-economic context of Africa.

3.4.1 Genetic diversity and effective population size

Overall genetic diversity (neutral and potentially adaptive combined), was in a similar range for both Tanganyika sardines. Heterozygosity was similar, but nucleotide diversity was slightly lower in *S. tanganicae* (H_o = 0.31, π = 0.22 %) than in *L. miodon* from Lake Tanganyika (H_o = 0.30, π = 0.31 %) and *L. miodon* from other lakes (H_o = 0.29

– 0.30, π = 0.26 – 0.32 %) (**Paper II & III**). The lower value in *S. tanganicae* could be caused by a higher amount of gene flow and homogenization in the absence of North-South adaptive variation. Genetic diversity measurements in other small pelagic clupeids are similar, but seem to depend strongly on the marker type. For example, based on whole genome sequencing data, the estimate for nucleotide diversity in *C. harengus* is π ≈ 0.30 % (Feng et al. 2017). Allozyme data of the overexploited *Sardinella aurita* suggested low heterozygosity (H_o = 0.01) (Chikhi et al. 1998), but a more recent study using RAD-seq showed much higher values (H_o = 0.31 – 0.32, π = 0.28 – 0.29 %) (Takyi 2019). In *Setipinna tenuifilis,* which has experienced a decline due to overfishing, a Hi-C genome combined with RAD-seq data showed much lower values (H_o = 0.15 – 0.18) (Liu et al. 2023). A study on mtDNA on the same species showed highly variable nucleotide diversity among populations (π = 0.2 – 0.7 %). Similar and slightly lower values were also found for *Sprattus sprattus* (RAD-seq, H_o = 0.25 – 0.29) (McKeown et al. 2020) and *Sardinella lemuru* (RAD-seq SNP-panel, H_o = 0.22 – 0.31) (Labrador et al. 2022). Much higher estimates are rare, except when measured with microsatellite data (e.g. *Alosa alosa* and *Alosa fallax*: H_o = 0.24 – 0.97) (Faria et al. 2004).

Based on nucleotide diversity, we were able to estimate contemporary N_e for the Tanganyika sardines. N_e was higher for *L. miodon* (N_e = 387,500) than for *S. tanganicae* (N_e = 275,000) and within species similar between the sub-basins and the five lakes (**Paper II & III**). To prevent the loss of genetic diversity, a minimum N_e of 500 is the recommended threshold (Allendorf et al. 1987; Hoban et al. 2020). Loss of heterozygosity due to acute reductions in N_e (e.g. population bottlenecks or small founder populations) is difficult to reverse, and larger N_e in subsequent generations does not bring back lost diversity (Allendorf et al. 1987). The introductions of *L. miodon* into the four other lakes did not seem to have a strong effect on N_e or genetic diversity, suggesting that the propagule size was sufficiently large. Based on genetic diversity in other small pelagic clupeids and our N_e estimates, we conclude that genetic diversity in the Tanganyika sardines is not critically low, perhaps thanks to their large effective population sizes in all populations (**Paper II & III**). Nevertheless, in the absence of

historical estimates we cannot yet exclude genetic erosion in Lake Tanganyika. To ensure a sustainable fishery, we recommend genetic diversity and effective population size analyses of historical samples to establish whether the values we observe are a norm or a minimum for these species. This endeavour has recently been initiated by an international team of scientists working on the historical evolution of the three main pelagic fisheries targets in Lake Tanganyika (https://site.nord.no/chipolata/). In addition, genetic erosion can be hidden over time, for example as recessive deleterious alleles, and only return to 'collect its debt' later (Bertorelle et al. 2022; Jackson et al. 2022).

3.4.2 Genetic population structure

In the 1999 LTR plan, the sardine stocks of both *S. tanganicae* and *L. miodon* were still defined as separate (northern and southern stocks). The results of our genetic population structure analyses (**Paper II**), together with those by De Keyzer et al. (2019) and Junker et al. (2020) suggest that this may be an appropriate view for *L. miodon,* but not for *S. tanganicae.* The former exhibits low genome-wide levels of differentiation but high levels at specific loci associated with an inversion and possibly adaptive loci, while the latter shows at most a weak pattern of IBD with enough migrants to maintain mixing. The precautionary principle in this case would mean to assume that northern and southern populations of *L. miodon* could be harmed and their lake-wide genetic diversity reduced by targeting the species as a single stock (Ryman et al. 1995). However, separate management at this stage would be premature, as the populations do not seem to be reproductively isolated. Such information should first be translated into practical guidelines by a team of (at least) one population geneticist, one fisheries scientist, and one fisheries manager. The latter two would preferentially already be involved in local fisheries management. The information subsequently needs to reach local communities, for example via the national points of contact of the LTA. Labelling wide-spread or common species as locally genetically distinct (e.g. 'unique', 'special', or 'local') may motivate people to feel more responsibility and take conservation action (Eyster et al. 2022). In addition, the amount

of gene flow needed to counter significant genetic differentiation, as seen in *S. tanganicae,* may mismatch with the amount of dispersal needed to replenish local numbers (Hauser and Carvalho 2008). This may not harm the stock genetically but could result in a shortage of fish available to the local community. To prevent this, the stock should be regularly monitored (e.g. through catch statistics) at different landing site aggregations as well as for the entire stock. Populations in other lakes are panmictic within each lake with the exception of Cahora Bassa (**Paper III,** see section 3.3.2). According to our results, the stocks of *L. miodon* in Lake Kivu, Itezhi-Tezhi, and Kariba can safely be monitored and managed on a lake-wide basis.

3.4.3 Potential for implementation

The large effective population sizes of the Tanganyika sardines (**Paper II & III**) are in agreement with the latest IUCN Red List assessment, where they were classified as 'least concern' (Ntakimazi 2006a, b). However, the indicators used for the IUCN Red List assessments are not well connected to genetic diversity (Hoban et al. 2020) and the assessment of *L. miodon* does not yet include populations in other lakes than Lake Tanganyika. We thus also evaluate our findings to the indicators proposed by Hoban et al. (2020). The first indicator is the number of populations within species with N_e above 500 versus those with N_e below 500. Assuming a single near-panmictic population in *S. tanganicae* (**Paper II**) and seven populations in *L. miodon* (one in each lake except Tanganyika and Cahora Bassa, where we assume two each, **Paper III**), 100% of the populations meet this criterium. The second indicator is the proportion of distinct populations maintained within species. Assuming the same populations as for the first indicator, no losses of distinct populations are known to date. As suggested in section 3.4.2, the putatively distinct populations should be monitored separately to ensure continued success for this indicator. The third indicator is the number of species and populations in which genetic diversity is being monitored using DNA-based methods. At present, not a single population of either of the two sardine species is being routinely monitored using DNA-based methods. The genetic and genomic studies conducted so far have been project-bound in both time and funding.

Even when genetic diversity is consistently monitored, what can practically be done with genetic indicators, specifically in the context of Africa? Surveys of 123 experts in various fish- and fisheries-related fields (**Paper IV**) show that there is a high degree of awareness and enthusiasm about genetic indicators in African countries, suggesting that there is potential for implementation. The most important hurdle mentioned by African respondents was a lack of resources and expertise. This was the most important hurdle for both non-genetic and genetic indicators. When resources for infrastructure are limited, creative and cost-effective solutions are needed. To ensure sustainability, a good ratio of sampling frequency and information returns is necessary (Plisnier et al. 2018). One option is to outsource library preparation and sequencing to avoid large investments in expensive instruments and the cost of keeping a trained technician to maintain them. Some sequencing centres offer a complete sample-to-data service that do not even require prior DNA extraction. The cheapest (per sample) method to measure genetic diversity and effective population size is high-throughput multiplexed shotgun libraries. A continuous genetic sampling program (ethanol preservation, in fridges or freezers if possible), routinely (e.g. every ten years) multiplexing samples from different years (e.g. sample individuals twice every ten years) and locations combined with outsourcing library prep and sequencing could be a feasible approach. Analysis can be performed in the free software R, which does not require access to a specialized server, from start to finish, and could be learned in a week-long workshop. An option to monitor the potentially adaptive variation and structure in *L. miodon* would be to introduce extremely simplified protocols. For example, development of a PCR- or restriction-based assay to genotype the inversion in *L. miodon* would allow monitoring of proportions of the karyotypes in the population with minimal equipment (small PCR machine and electrophoresis system, e.g. https://theminione.com/) and training. Sharing one laboratory for analysis of samples from the entire lake would also reduce costs and the number of trainings (Plisnier et al. 2018). While the enthusiasm and possibilities for genetic indicator application are encouraging, genetic data for most species is still lacking (Hoban et al. 2021) and a shift

will not be fast. In the meantime, non-genetic proxies can be used to report on genetic parameters, for example census population size, derived from fishery statistics, as a proxy for effective population size (Hoban et al. 2022).

Research on the African Great Lakes is often short-term, not standardized, local, and conducted by researchers outside of the region (Plisnier et al. 2018; Obiero et al. 2020). Similarly, education and professional development opportunities are rare, inconsistent, and limited by governmental funding and harsh competition for funded opportunities (Müller et al. 2015; Obiero et al. 2020; O'Connell et al. 2022). The LTA, aided by the FAO, is the leading organization for continued learning opportunities around Lake Tanganyika and emphasizes capacity building by its working groups (Reynolds 1999). For example, they have organized workshops around aquaculture (Bodiguel and Swan 2014), awareness of sustainable fisheries practices for local fishing communities (LTA 2014), and climate change in the African lakes (LTA 2010). The African Center for Aquatic Research and Education (ACARE) is a recently established organization and contains a network of researchers around the African Great Lakes. They also have a partnership with the International Institute for Sustainable Development (IISD), which provides legal and policy expertise (https://www.agl-acare.org/). One of their key goals is to enable low-cost professional development e.g. through webinars, online resources, and conference workshops. Obiero et al. (2020) gives an overview of existing international initiatives around the Great Lakes that focus on enhancing both higher education and professional development opportunities. ACARE aims to connect such initiatives and facilitate information and (human) resource exchange between them. However, none of them are in the close vicinity of Lake Tanganyika and only one is in a riparian country.

4 Conclusion

This thesis has addressed critical challenges in the fisheries management in the Great Lakes Region. We focused on Lake Tanganyika's two dominant pelagic fisheries targets, *S. tanganicae* and *L. miodon* through a genetic and socio-economic lens. Using modern genomic methods, we have generated high quality (mito)genomic resources for both species and characterized the ecological and genomic factors that have shaped their divergence and contemporary evolution. On a macro-evolutionary scale, historical allopatry within the Malagarasi-Congo river system, ecological opportunity during the onset of lacustrine conditions in the proto-Lake Tanganyika, and sex chromosome turnover have emerged as the most important factors. On a micro-evolutionary scale, long-distance migration and gene flow cause the near-panmictic structure in *S. tanganicae*, while inversion-facilitated divergence in vision and sperm property related genes, and historical (within-lake) and contemporary (within and between-lake) allopatry have shaped the more prominent population structure of *L. miodon.* Using the same genomic resources, we estimated management-relevant genetic indicators, specifically genetic diversity, population structure, and effective population size. Genetic diversity and effective population size were consistently high in all populations. Population structure was shallow, except for strong divergence in sex- and inversion-linked regions. We integrated these findings in a social and policy context by exploring the applicability of genetic indicators in African countries. We mapped the diverse perspectives on this topic with a comparative analysis of African and non-African fish and fisheries experts. This analysis showed a high awareness and willingness, but low current implementation of genetic indicators in Africa, and identified a lack of resources and expertise as the number one hurdle. Together, the chapters of this thesis provide essential resources for the future genetic monitoring of the Tanganyika sardines and related species, highlight mechanisms that shape the evolution of small pelagic fishes, and present recommendations for fisheries management of these important fisheries targets and African fishes in general.

References

Abdulkarim B, Bwathondi POJ, Benno B (2014) Comparative evaluation of the proximate and mineral compositions of *Stolothrissa tanganicae* and *Limnothrissa miodon* from Lake Tanganyika. J Rev 2:348–354. https://doi.org/10.14196/SJR.V2I12.1120

Adams PB (1980) Life history patterns in marine fishes and their consequences for fisheries management. Fish Bull 78:1–12

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11:697–709. https://doi.org/10.1038/nrg2844

Allendorf FW, Lundquist LL (2003) Introduction: Population Biology, Evolution, and Control of Invasive Species. Conserv Biol 17:24–30. https://doi.org/https://doi.org/10.1046/j.1523-1739.2003.02365.x

Allendorf FW, Ryman N, Utter FM (1987) Genetics and fishery management. In: Ryman N, Utter F (eds) Population Genetics and Fishery Management. pp 1–20

Anseeuw D, Raeymaekers JAM, Busselen P, Verheyen E, Snoeks J (2011). Low genetic and morphometric intraspecific divergence in peripheral *Copadichromis* populations (Perciformes: Cichlidae) in the Lake Malawi basin. Int J Evol Biol 2001:835946. https://doi.org/10.4061/2011/835946

Anseeuw D (2014) Population substructuring in the Utaka (Teleostei, Cichlidae) from Lake Malawi (East Africa): do genetic and (eco-)morphological differentiation go hand in hand? [PhD thesis] KULeuven, Belgium

Bakker VJ, Baum JK, Brodie JF, Salomon A, Dickson B, Gibbs H, et al (2010) The changing landscape of conservation science funding in the United States. Conserv Lett 3:435–444. https://doi.org/10.1111/j.1755-263X.2010.00125.x

Baltazar-Soares M, Hinrichsen HH, Eizaguirre C (2018) Integrating population genomics and biophysical models towards evolutionary-based fisheries management. ICES J Mar Sci 75:1245–1257. https://doi.org/10.1093/icesjms/fsx244

Barrio AM, Lamichhaney S, Fan G, Rafati N, Pettersson M, Zhang H (2016) The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. Elife 5:e12081. https://doi.org/10.7554/eLife.12081

Barth JMI, Berg PR, Jonsson PR, Bonanomi S, Corell H, Hemmer-Hansen J, et al (2017) Genome architecture enables local adaptation of Atlantic cod despite high connectivity. Mol Ecol 26:4452–4466. https://doi.org/10.1111/mec.14207

Begg GA, Waldman JR (1999) An holistic approach to fish stock identification. Fish Res 43:35–44. https://doi.org/https://doi.org/10.1016/S0165-7836(99)00065-X

Bell-Cross G, Bell-Cross B (1971) Introduction of *Limnothrissa miodon* and *Limnocaridina tanganicae* from Lake Tanganyika into Lake Kariba. Fish Res Bull Zambia 5:207–214

Berg PR, Jentoft S, Star B, Ring K, Knutsen H, Lien S, et al (2015) Adaptation to Low Salinity Promotes Genomic Divergence in Atlantic Cod (*Gadus morhua* L.). Genome Biol Evol 7:1644–1663. https://doi.org/10.1093/gbe/evv093

Bernacsek GM, Lopes S (1984) Cahora Bassa (Mozambique). CIFA Technical Paper 10:21–24

Bernatchez L, Wellenreuther M, Araneda C, Ashton DT, Barth JMI, Beacham, TD, et al (2017) Harnessing the Power of Genomics to Secure the Future of Seafood. Trends Ecol Evol 32:665–680. https://doi.org/10.1016/j.tree.2017.06.010

Bernery C, Bellard CA, Courchamp F, Brosse S, Gozlan R, Jarić I, et al (2022) Freshwater Fish Invasions: A Comprehensive Review. Annu Rev Ecol Evol Syst 53:57. https://doi.org/10.1146/annurev-ecolsys-032522-015551

Bertorelle G, Raffini F, Bosse M, Bortoluzzi C, Iannucci A, Trucchi E, et al (2022) Genetic load: genomic estimates and applications in non-model animals. Nat Rev Genet 23:492–503. https://doi.org/10.1038/s41576-022-00448-x

Beverton R (1990) Small marine pelagic fish and the threat of fishing; are they endangered? J Fish Biol 37:5–16. https://doi.org/10.1111/j.1095-8649.1990.tb05015.x

Bieniek-Kobuszewska M, Borkowski J, Panasiewicz G, Nowakowski JJ (2020) Impact of conservation and hunting on big game species: comparison of the genetic diversity of the red deer population groups from a national park and neighboring hunting areas in northern Poland. Eur Zool J 87:603–615. https://doi.org/10.1080/24750263.2020.1822936

Blache J (1964) Les poissons du bassin du Tchad et du bassin adjacent du Mayo Kebbi. Étude systématique et biologique. Mémoires de l'Office de la Recherche Scientifique et Technique Outre-Mer(4)

Bloom DD, Egan JP (2018) Systematics of clupeiformes and testing for ecological limits on species richness in a trans-marine/freshwater clade. Neotrop Ichthyol 16:492–503. https://doi.org/10.1590/1982-0224-20180095

Bloom DD, Lovejoy NR (2014) The evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes (herring and allies). Proc R Soc B: Biol Sci 281:20132081. https://doi.org/10.1098/rspb.2013.2081

Bodiguel C, Swan J (2014) Report of the Lake Tanganyika Authority workshop to develop a protocol on aquaculture. Ebene, Mauritius

Boulenger GA (1898) Catalogue of the fresh-water fishes of Africa. British Museum (Natural History), Taylor and Francis, London, UK

Brant SV, Loker ES (2009) Molecular systematics of the avian schistosome genus *Trichobilharzia* (Trematoda: Schistosomatidae) in North America. J Parasitol 95:941–963. https://doi.org/10.1645/GE-1870.1

Brown KJ, Rüber L, Bills R, Day JJ (2010) Mastacembelid eels support Lake Tanganyika as an evolutionary hotspot of diversification. BMC Evol Biol 10:188. https://doi.org/10.1186/1471-2148-10-188

Brugha R (2000) Stakeholder analysis: a review. Health Policy Plan 15:239–246. https://doi.org/10.1093/heapol/15.3.239

Canales-Aguirre CB, Ferrada-Fuentes S, Galleguillos R, Hernández CE (2016) Genetic Structure in a Small Pelagic Fish Coincides with a Marine Protected Area: Seascape Genetics in Patagonian Fjords. PLoS One 11:e0160670. https://doi.org/10.1371/journal.pone.0160670

Cane MA, Molnar P (2001) Closing of the Indonesian seaway as a precursor to east African aridification around 3–4 million years ago. Nature 411:157–162. https://doi.org/10.1038/35075500

Cassemiro FAS, Albert JS, Antonelli A, Menegotto A, Wüest R, Cerezer F, et al (2023) Landscape dynamics and diversification of the megadiverse South American freshwater fish fauna. PNAS 120:e2211974120. https://doi.org/10.1073/pnas.2211974120

CBD (2011) Text and Annexes. Secretariat for the Convention on Biological Diversity. Montréal, Canada

CBD (2020) Global Biodiversity Outlook 5. Montréal, Canada

Chan KMA, Balvanera P, Benessaiah K, Chapman M, Díaz S, Gómez-Baggethun E, et al (2016) Why protect nature? Rethinking values and the environment. PNAS 113:1462–1465. https://doi.org/10.1073/pnas.1525002113

Charlesworth B (2009) Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10:195–205. https://doi.org/10.1038/nrg2526

Chen RS, Soulsbury CD, Lebigre C, Ludwig G, van Oers K, Hoffman J (2023) Effects of hunting on genetic diversity, inbreeding and dispersal in Finnish black grouse (*Lyrurus tetrix*). Evol Appl 16:625–637. https://doi.org/10.1111/eva.13521

Chikhi L, Bonhomme F, Jean-François A (1998) Low genetic variability in a widely distributed and abundant clupeid species, *Sardinella aurita*: new empirical results and interpretations. J Fish Biol 52:861–878. https://doi.org/10.1111/j.1095-8649.1998.tb00588.x

Cioffi MB, Liehr T, Trifonov V, Molina W, Bertollo L, et al (2013) Independent sex chromosome evolution in lower vertebrates: A molecular cytogenetic overview in the Erythrinidae fish family. Cytogenet Genome Res 141:186–194. https://doi.org/10.1159/000354039

Cohen AS, Soreghan MJ, Scholz CA (1993) Estimating the age of formation of lakes: an example from Lake Tanganyika, East African Rift system. Geology 21:511–514. https://doi.org/10.1130/0091-7613(1993)021<0511:ETAOFO>2.3.CO;2

Collart A (1960) L'introduction du *Stolothrissa tanganicace* (Ndagala) au Lac Kivu. Bulletin Agricole du Congo Belge 51:975–985

Connallon T, Olito C, Dutoit L, Papoli H, Ruzicka F, Yong L (2018) Local adaptation and the evolution of inversions on sex chromosomes and autosomes. Philos Trans R Soc Lond B Biol Sci 373:20170423. https://doi.org/10.1098/rstb.2017.0423

Coulter GW (1991) Lake Tanganyika and its life. Oxford University Press, UK

Crutsinger GM (2016) A community genetics perspective: Opportunities for the coming decade. New Phytol 210:65–70. https://doi.org/10.1111/nph.13537

Cummings MP, Otto SP, Wakeley J (1995) Sampling properties of DNA sequence data in phylogenetic analysis. Mol Biol Evol 12:814–822. https://doi.org/10.1093/oxfordjournals.molbev.a040258

Cury P (2000) Small pelagics in upwelling systems: patterns of interaction and structural changes in “wasp-waist” ecosystems. ICES J Mar Sci 57:603–618. https://doi.org/10.1006/jmsc.2000.0712
Day JJ, Bills R, Friel JP (2009) Lacustrine radiations in African *Synodontis* catfish. J Evol Biol 22:805–817. https://doi.org/10.1111/j.1420-9101.2009.01691.x
Day JJ, Wilkinson M (2006) On the origin of the *Synodontis* catfish species flock from Lake Tanganyika. Biol Lett 2:548–552. https://doi.org/10.1098/rsbl.2006.0532
De Iongh HH, Spliethoff PC, Frank VG (1983) Feeding habits of the clupeid *Limnothrissa miodon* (Boulenger), in Lake Kivu. Hydrobiologia 102:113–122. https://doi.org/10.1007/BF00006074
De Keyzer ELR, De Corte Z, Van Steenberge M, Raeymaekers JAM, Calboli FCF, Kmentová N, et al (2019) First genomic study on Lake Tanganyika sprat *Stolothrissa tanganicae*: a lack of population structure calls for integrated management of this important fisheries target species. BMC Evol Biol 19:1–15. https://doi.org/10.1186/s12862-018-1325-8
De Keyzer ELR, Masilya Mulungula P, Alunga Lufungula G, Amisi Manala C, Andema Muniali A, Bashengezi Cibuhira P, et al (2020) Local perceptions on the state of the pelagic fisheries and fisheries management in Uvira, Lake Tanganyika, DR Congo. J Great Lakes Res 46:1740–1753. https://doi.org/10.1016/j.jglr.2020.09.003
de Sousa RPC, Silva-Oliveira GC, Furo IO, de Oliveira-Filho A, de Brito C, Rabelo L, et al (2021) The role of the chromosomal rearrangements in the evolution and speciation of Elopiformes fishes (Teleostei; Elopomorpha). Zool Anz 290:40–48. https://doi.org/10.1016/j.jcz.2020.11.003
Decru E, Vranken N, Maetens H, Mertens De Vry A, Kayenbergh A, Snoeks J, et al (2022) DNA barcoding the Lake Edward basin: high taxonomic coverage of a tropical freshwater ichthyofauna. Hydrobiologia 849:1743–1762. https://doi.org/10.1007/s10750-022-04812-0
Devlin RH, Nagahama Y (2002) Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208:191–364. https://doi.org/10.1016/S0044-8486(02)00057-1
Dichmont CM, Ovenden JR, Berry O, Welch D, Buckworth R (2012) Scoping current and future genetic tools, their limitations and their applications for wild fisheries management. FRDC 2011/035
Duchêne S, Archer FI, Vilstrup J, Caballero S, Morin P (2011) Mitogenome phylogenetics: The impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. PLoS One 6:e27138. https://doi.org/10.1371/journal.pone.0027138
Dufresnes C, Crochet PA (2022) Sex chromosomes as supergenes of speciation: Why amphibians defy the rules? Philos Trans R Soc Lond B Biol Sci 377:20210202. https://doi.org/10.1098/rstb.2021.0202

Dufresnes C, Litvinchuk SN, Rozenblut-Kościsty B, Rodrigues N, Perrin N, Crochet P, et al (2020) Hybridization and introgression between toads with different sex chromosome systems. Evol Lett 4:444–456. https://doi.org/10.1002/evl3.191

Dumont HJ (1986) The Tanganyika Sardine in Lake Kivu: Another Ecodisaster for Africa? Environ Conserv 13:143–148. https://doi.org/10.1017/S0376892900036742

Eccles DH (1985) Lake Flies and Sardines – A Cautionary Note. Biol Conserv 33:309–333. https://doi.org/10.1016/0006-3207(85)90074-6

Egan JP, Bloom DD, Kuo CH, Hammer M, Tongnunui P, Iglésias S, et al (2018) Phylogenetic analysis of trophic niche evolution reveals a latitudinal herbivory gradient in Clupeoidei (herrings, anchovies, and allies). Mol Phylogenet Evol 124:151–161. https://doi.org/10.1016/j.ympev.2018.03.011

El Taher A, Ronco F, Matschiner M, Salzburger W, Böhner A (2021) Dynamics of sex chromosome evolution in a rapid radiation of cichlid fishes. Sci Adv 7:eabe8215. https://doi.org/10.1126/sciadv.abe8215

European Environment Agency (2023) FAO Fishery Glossary. https://www.eea.europa.eu/help/glossary/eea-glossary/fish-stock. Accessed 4 August 2024

Eyster HN, Olmsted P, Naidoo R, Chan KMA (2022) Motivating conservation even for widespread species using genetic uniqueness and relational values. Biol Conserv 266:109438. https://doi.org/10.1016/j.biocon.2021.109438

Ezaz T, Stiglec R, Veyrunes F, Marshall Graves JA (2006) Relationships between Vertebrate ZW and XY Sex Chromosome Systems. Current Biology 16:R736–R743. https://doi.org/10.1016/j.cub.2006.08.021

FAO (2019) The State of the World’s Aquatic Genetic Resources for Food and Agriculture. Rome, Italy

FAO (2022) The State of World Fisheries and Aquaculture. Rome, Italy. https://doi.org/10.4060/cc0461en

FAO (2023) Illuminating hidden harvests: The contributions of small-scale fisheries to sustainable development. https://doi.org/https://doi.org/10.4060/cc4576en

FAO (1995) Code of conduct for responsible fisheries. Rome, Italy

FAO (2003) The convention on the sustainable management of Lake Tanganyika

Faria R, Wallner B, Weiss S, Alexandrino P (2004) Isolation and characterization of eight dinucleotide microsatellite loci from two closely related clupeid species (*Alosa alosa* and *A. fallax*). Mol Ecol Notes 4:586–588. https://doi.org/10.1111/j.1471-8286.2004.00745.x

Feder JL, Nosil P, Flaxman SM (2014) Assessing when chromosomal rearrangements affect the dynamics of speciation: Implications from computer simulations. Front Genet 5:1–14. https://doi.org/10.3389/fgene.2014.00295

Feng C, Pettersson M, Lamichhaney S, Rubin C, Rafati N, Casini M, et al (2017) Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. Elife 6:e23907. https://doi.org/10.7554/eLife.23907

Frankham R (2005) Resolving the genetic paradox in invasive species. Heredity 94:385–385. https://doi.org/10.1038/sj.hdy.6800634

Frankham R (2010) Challenges and opportunities of genetic approaches to biological conservation. Biol Conserv 143:1919–1927. https://doi.org/10.1016/j.biocon.2010.05.011

Froese R, Pauly D (2024) FishBase. www.fishbase.org. Accessed 11 Aug 2024

Fryer G (2006) Evolution in ancient lakes: radiation of Tanganyikan atyid prawns and speciation of pelagic cichlid fishes in Lake Malawi. Hydrobiologia 568:131–142. https://doi.org/10.1007/s10750-006-0322-x

Furman BLS, Metzger DCH, Darolti I, Wright A, Sandkam B, Almeida P, et al (2020) Sex Chromosome Evolution: So Many Exceptions to the Rules. Genome Biol Evol 12:750–763. https://doi.org/10.1093/gbe/evaa081

Gagnaire P-A, Gaggiotti OE (2016) Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. Curr Zool 62:603–616. https://doi.org/10.1093/cz/zow088

Gaither MR, Bowen BW, Rocha LA, Briggs JC (2016) Fishes that rule the world: circumtropical distributions revisited. Fish Fish 17:664–679. https://doi.org/10.1111/faf.12136

Gaunt MW, Miles MA (2002) An Insect Molecular Clock Dates the Origin of the Insects and Accords with Palaeontological and Biogeographic Landmarks. Mol Biol Evol 19:748–761. https://doi.org/10.1093/oxfordjournals.molbev.a004133

Getahun A, Wakjira M, Nyingi DW (2020) Social, economic and management status of small-scale fisheries in Omo River Delta and Ethiopian side of Lake Turkana, southern Ethiopia. Ecohydrol Hydrobiol 20:323–332. https://doi.org/10.1016/j.ecohyd.2020.05.008

Glover KA, Skaala Ø, Limborg M, Kvamme C, Torstensen E (2011) Microsatellite DNA reveals population genetic differentiation among sprat (*Sprattus sprattus*) sampled throughout the Northeast Atlantic, including Norwegian fjords. ICES J Mar Sci 68:2145–2151. https://doi.org/10.1093/icesjms/fsr153

Grant WS, Bowen BW (1998) Shallow Population Histories in Deep Evolutionary Lineages of Marine Fishes: Insights From Sardines and Anchovies and Lessons for Conservation. The American Genetic Association, USA

Haq BU, Hardenbohl J, Vail PR (1987) Chronology of fluctuating sea levels since the Triassic (250 million years ago to present). Science 235:1156–1167. https://doi.org/10.1126/science.235.4793.1156

Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. Fish Fish 9:333–362. https://doi.org/10.1111/j.1467-2979.2008.00299.x

Henning F, Trifonov V, Ferguson-Smith MA, de Almeida-Toledo LF (2008) Non-homologous sex chromosomes in two species of the genus *Eigenmannia* (Teleostei: Gymnotiformes). Cytogenet Genome Res 121:55–58. https://doi.org/10.1159/000124382

Hill WG, Zhang XS (2012) On the pleiotropic structure of the genotype-phenotype map and the evolvability of complex organisms. Genetics 190:1131–1137. https://doi.org/10.1534/genetics.111.135681

Hoban S, Bruford M, D'Urban Jackson J, Lopes-Fernandes M, Heuertz M, Hohenlohe P, et al (2020) Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. Biol Conserv 248:108654. https://doi.org/10.1016/j.biocon.2020.108654

Hoban S, Campbell CD, da Silva JM, Ekblom R, Funk WC, Garner BA, et al (2021). Genetic diversity is considered important but interpreted narrowly in country reports to the Convention on Biological Diversity: Current actions and indicators are insufficient. Biol Conserv 261:109233. https://doi.org/10.1016/j.biocon.2021.109233

Hoban S, Da Silva J, Mastretta-Yanes A, Grueber C, Heuertz M, Hunter M, et al (2022) Monitoring status and trends in genetic diversity for the Convention on Biological Diversity: an ongoing assessment of genetic indicators in nine countries. Conserv Letters 16:e12953. https://doi.org/10.22541/au.166627692.27077414/v1

Hoffmann AA, Rieseberg LH (2008) Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? Annu Rev Ecol Evol Syst 39:21–42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: Implications for landscape genetics. Landsc Ecol 21:797–807. https://doi.org/10.1007/s10980-005-5245-9

Hooper DM, Price TD (2017) Chromosomal inversion differences correlate with range overlap in passerine birds. Nat Ecol Evol 1:1526–1534. https://doi.org/10.1038/s41559-017-0284-6

Hubert N, Hanner R (2015) DNA Barcoding, species delineation and taxonomy: a historical perspective. DNA Barcodes 3:44–58. https://doi.org/10.1515/dna-2015-0006ï

Ivory SJ, Blome MW, King JW, McGlue M, Cole J, Cohen A (2016) Environmental change explains cichlid adaptive radiation at Lake Malawi over the past 1.2 million years. PNAS 113:11895–11900. https://doi.org/10.1073/pnas.1611028113

Jackson HA, Percival-Alwyn L, Ryan C, Albeshr M, Venturi L, Morales H, et al (2022) Genomic erosion in a demographically recovered bird species during conservation rescue. Conserv Biol 36:. https://doi.org/10.1111/cobi.13918

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al (2012) The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61. https://doi.org/10.1038/nature10944

Junker J, Rick JA, McIntyre PB, Kimirei I, Sweke E, Mosille J, et al (2020) Structural genomic variation leads to genetic differentiation in Lake Tanganyika's sardines. Mol Ecol 29:3277–3298. https://doi.org/10.1111/mec.15559

Junor FJR, Begg GW (1971) A note on the successful introduction of *Limnothrissa miodon* (Boulenger) "the Lake Tanganyika sardine" to Lake Kariba. J Limnol Soc South Afr 16:8–14

Kamenova S, Bartley TJ, Bohan DA, Boutain J, Colautti R, Domaizon I, et al (2017) Invasions Toolkit: Current Methods for Tracking the Spread and Impact of Invasive Species. Adv Ecol Res 56:85–182. https://doi.org/10.1016/bs.aecr.2016.10.009

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, UK

Kirkpatrick M (2010) How and why chromosome inversions evolve. PLoS Biol 8:e1000501. https://doi.org/10.1371/journal.pbio.1000501

Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. Genetics 173:419–434. https://doi.org/10.1534/genetics.105.047985

Kitano J, Peichel CL (2012) Turnover of sex chromosomes and speciation in fishes. Environ Biol Fishes 94:549–558. https://doi.org/10.1007/s10641-011-9853-8

Kmentová, N, Cruz-Laufer AJ, Milec LJM, Moons T, Heeren S, van den Hoorn E, et al (2024) Host lifestyle and parasite interspecific facilitation mediate co-infection in a species-poor host–parasite system. Oikos (online): e10360. https://doi.org/10.1111/oik.10360

Kmentová N, Thys K, Hahn C, Vorel J, Koblmüller S, Van Steenberge M, et al (2023) Comparative population mitogenomics of fish parasites reveals contrasting geographic pattern in the pelagic zone of Lake Tanganyika. Authorea. https://doi.org/10.22541/au.167274604.41657501/v1

Kmentová N, Van Steenberge M, Thys van den Audenaerde DFE, Nhiwatiwa T, Muterezi Bukinga F, Mulimbwa NT, et al (2019) Co-introduction success of monogeneans infecting the fisheries target *Limnothrissa miodon* differs between two non-native areas: the potential of parasites as a tag for introduction pathway. Biol Invasions 21:757–773. https://doi.org/10.1007/s10530-018-1856-3

Koblmüller S, Schöggl CA, Lorber CJ, Van Steenberge M, Kmentová N, Vanhove MPM, et al (2021) African lates perches (Teleostei, Latidae, *Lates*): Paraphyly of Nile perch and recent colonization of Lake Tanganyika. Mol Phylogenet Evol 160:. https://doi.org/10.1016/j.ympev.2021.107141

Koblmüller S, Zangl L, Börger C, Daill D, Vanhove MPM, Strumbauer C, et al (2019) Only true pelagics mix: comparative phylogeography of deepwater bathybatine cichlids from Lake Tanganyika. Hydrobiologia 832:93–103. https://doi.org/10.1007/s10750-018-3752-3

Kolding J, van Zwieten P, Marttin F, Funge-Smith S, Poulain F (2019) Freshwater small pelagic fish and their fisheries in major African lakes and reservoirs in relation to food security and nutrition. FAO Fisheries and Aquaculture Technical Paper 642. Rome, Italy

Korpelainen H (1990) Sex ratios and conditions required for environmental sex determination in animals. Biological Reviews 65:147–184. https://doi.org/10.1111/j.1469-185X.1990.tb01187.x

Koskela J, Lefèvre F, Schueler S, Kraigher H, Olrik D, Hubert J, et al (2013) Translating conservation genetics into management: Pan-European minimum requirements for dynamic conservation units of forest tree genetic diversity. Biol Conserv 157:39–49. https://doi.org/10.1016/j.biocon.2012.07.023

Kuusipalo L (1999) Genetic variation in the populations of pelagic clupeids *Stolothrissa tanganicae* and *Limnothrissa miodon* and nile perch (*Lates stappersii*, *L. mariae*) in Lake Tanganyika. FAO/Finnida Research for the management of the Fisheries of Lake Tanganyika 92:28

Labrador K, Palermo JD, Agmata A, Ravago-Gotanco R, Pante M (2022) Restriction Site-Associated DNA Sequencing Reveals Local Adaptation Despite High Levels of Gene Flow in *Sardinella lemuru* (Bleeker, 1853) along the Northern Coast of Mindanao, Philippines. Front Mar Sci 9:. https://doi.org/10.3389/fmars.2022.766936

Lahn BT, Page DC (1999) Four Evolutionary Strata on the Human X Chromosome. Science 286:964–967. https://doi.org/10.1126/science.286.5441.964

Laikre L, Hoban S, Bruford MW, Segelbacher G, Allendorf F, Gajardo G, et al (2020) Post-2020 goals overlook genetic diversity. Sci Let 367:1083–1085. https://doi.org/10.1126/science.abb2748

Lavoué S (2020) Origins of Afrotropical freshwater fishes. Zool J Linn Soc 188:345–411. https://doi.org/10.1093/zoolinnean/zlz039

Lavoué S, Miya M, Moritz T, Nishida M (2012) A molecular timescale for the evolution of the African freshwater fish family Kneriidae (Teleostei: Gonorynchiformes). Ichthyol Res 59:104–112. https://doi.org/10.1007/s10228-011-0258-7

Lavoué S, Miya M, Musikasinthorn P, Chen W, Nishida M (2013) Mitogenomic Evidence for an Indo-West Pacific Origin of the Clupeoidei (Teleostei: Clupeiformes). PLoS One 8:e56485. https://doi.org/10.1371/journal.pone.0056485

Lévêque C (1996) Introduction of fish species in freshwaters: a major threat to aquatic biodiversity? In: Di Castri F, Younès T (eds) Biodiversity, science and development: towards a new partnership. Cab International. pp 446–451

Lévêque C, Paugy D, Teugels G (1990) The fresh and brackish water fishes of West Africa, volume 1. Tervuren, Belgium

Lindholm A, Breden F (2002) Sex Chromosomes and Sexual Selection in Poeciliid Fishes. Am Nat 160:S214–S224. https://doi.org/10.1086/342898

Liu B, Li J, Peng Y, Zhang K, Liu Q, Jin X, et al (2023) Chromosome-level genome assembly and population genomic analysis reveal evolution and local adaptation in common hairfin anchovy (*Setipinna tenuifilis*). Mol Ecol. https://doi.org/10.1111/mec.17067

Liu B-J, Zhang K, Zhang S-F, Liu Y, Li J, Peng Y, et al (2022) Chromosome-level genome assembly of the dotted gizzard shad (*Konosirus punctatus*) provides insights into its adaptive evolution. Zool Res 43:217–220. https://doi.org/10.24272/j.issn.2095-8137.2021.351

Lowe-Mcconnell RH (1993) Fish Faunas of the African Great Lakes: Origins, Diversity, and Vulnerability. Conserv Biol 7:634–643. https://doi.org/10.1046/j.1523-1739.1993.07030634.x

Lowry DB, Hoban S, Kelley J, Lotterhos K, Reed L, Antolin M, et al (2017) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol Ecol Res 17:142–152. https://doi.org/10.1111/1755-0998.12635

LTA (2009) Strategic Action Programme. http://lta.iwlearn.org/management-program. Accessed 6 Sep 2023

LTA (2014) Awareness workshop on sustainable management of fisheries for fishing communities in Burundi. Rumonge, Burundi

LTA (2010) Regional Workshop on Climate Change and Fisheries in the African Great Lakes. Nairobi, Kenya

Lu G, Luo M (2020) Genomes of major fishes in world fisheries and aquaculture: Status, application and perspective. Aquac Fish 5:163–173. https://doi.org/10.1016/j.aaf.2020.05.004

Luikart G, Ryman N, Tallmon DA, Schwartz M, Allendorf F (2010) Estimation of census and effective population sizes: The increasing usefulness of DNA-based approaches. Conserv Genet 11:355–373. https://doi.org/10.1007/s10592-010-0050-7

Ma F, Wang Y, Su B, et al (2023) Gap-free genome assembly of anadromous *Coilia nasus*. Sci Data 10:360. https://doi.org/10.1038/s41597-023-02278-w

MacLean JA, Evans DO (1981) The Stock Concept, Discreteness of Fish Stocks, and Fisheries Management. Can J Fish Aquat Sci 38:1889–1898. https://doi.org/10.1139/f81-235

Mank JE, Avise JC (2006) The evolution of reproductive and genomic diversity in ray-finned fishes: insights from phylogeny and comparative analysis. J Fish Biol 69:1–27. https://doi.org/10.1111/j.1095-8649.2006.01132.x

Mank JE, Promislow DEL, Avise JC (2006) Evolution of alternative sex-determining mechanisms in teleost fishes. Biol J Linn Soc 87:83–93. https://doi.org/https://doi.org/10.1111/j.1095-8312.2006.00558.x

Mannini P, Aro E, Katonda I, Kassaka B, Mambona C, Milindi G, et al (1996) Pelagic fish stocks of Lake Tanganyika: Biology and exploitation. FAO/FINNIDA Research for the Management of Fisheries on Lake Tanganyika

Marshall (1991) The Impact of the Introduced Sardine *Limnothrissa miodon* on the Ecology of Lake Kariba. Biol Conserv 55:151–165. https://doi.org/10.1016/0006-3207(91)90054-D

Marshall BE (1984) Small pelagic fishes and fisheries in African inland waters. CIFA Technical Paper1 14:1–25

Marshall BE (1988) A preliminary assessment of the biomass of the pelagic sardine *Limnothrissa miodon* in Lake Kariba. J Fish Biol 32:515–524. https://doi.org/10.1111/j.1095-8649.1988.tb05391.x

Marshall BE (1995) Why is *Limnothrissa miodon* such a successful introduced species and is there anywhere else we should put it? In: Pitcher TJ, Hart PJB (eds) The Impact of Species Changes in African Lakes (Fish & Fisheries Series 18). Chapman & Hall, pp 527–545

Matthes H (1967) Preliminary investigations into the biology of the Lake Tanganyika Clupeidae. Fish Res Bull Zambia 4:1965–1966

McKeown NJ, Carpi P, Silva JF, Healey A, Shaw P, van der Kooij J (2020) Genetic population structure and tools for the management of European sprat (*Sprattus sprattus*). ICES J Mar Sci 77:2134–2143. https://doi.org/10.1093/icesjms/fsaa113

Miller DJ (1989) Introductions and Extinction of Fish in the African Great Lakes. Tree 4:56–59. https://doi.org/10.1016/0169-5347(89)90145-6

Miura I, Ohtani H, Nakamura M, Ichikawa Y, Saitoh K (1998) The Origin and Differentiation of the Heteromorphic Sex Chromosomes Z, W, X, and Y in the Frog *Rana rugosa*, Inferred from the Sequences of a Sex-Linked Gene, ADP/ATP Translocase. Mol Biol Evol 15:1612–1619. https://doi.org/10.1093/oxfordjournals.molbev.a025889

Molina WF, da Costa GWWF, de Bello Cioffi M, Bertollo LAC (2012) Chromosomal differentiation and speciation in sister-species of Grammatidae (Perciformes) from the Western Atlantic. Helgol Mar Res 66:363–370. https://doi.org/10.1007/s10152-011-0276-x

Mölsä H, Reynolds JE, Coenen EJ, Lindqvist O V. (1999) Fisheries research towards resource management on Lake Tanganyika. Hydrobiologia 407:1–24. https://doi.org/10.1023/A:1003712708969

Mora C, Myers RA, Coll M, Libralato S, Pitcher T, Sumaila R, et al (2009) Management effectiveness of the world’s marine fisheries. PLoS Biol 7:e1000131. https://doi.org/10.1371/journal.pbio.1000131

MSC (2021) MSC Briefing 01: Small pelagic fisheries - meeting future challenges

Mubamba R (1992) Introduction of Lake Tanganyika sardines into Itezhi-Tezhi reservoir, Zambia. In: Marshall BE, Mubamba R (eds) Papers presented at the symposium on biology, stock assessment and exploitation of small pelagic fish species in the African Great Lakes region, Bujumbura, Burundi

Muderhwa N, Matabaro L (2010) The introduction of the endemic fish species, *Lamprichthys tanganicanus* (Poeciliidae), from Lake Tanganyika into Lake Kivu: Possible causes and effects. Aquat Ecosyst Health Manag 13:203–213. https://doi.org/10.1080/14634981003800733

Mulimbwa N, Sarvala J, Raeymaekers JAM (2014) Reproductive activities of two zooplanktivorous clupeid fish in relation to the seasonal abundance of copepod prey in the northern end of lake Tanganyika. Belg J Zool 144:77–92. https://doi.org/10.1080/14634988.2014.883896

Mulimbwa N, Shirakihara K (1994) Growth, recruitment and reproduction of sardines (*Stolothrissa tanganicae* and *Limnothrissa miodon*) in northwestern Lake Tanganyika. Tropics 4:57–67. https://doi.org/10.3759/tropics.4.57

Mulimbwa NT, Milec LJM, Raeymaekers JAM, Sarvala J, Plisnier P-D, Marwa B, et al (2022) Spatial and seasonal variation in reproductive indices of the clupeids *Limnothrissa miodon* and *Stolothrissa tanganicae* in the Congolese waters of northern Lake Tanganyika. Belg J Zool 15:13–31. https://doi.org/10.26496/bjz.2022.96

Müller E, Appleton MR, Ricci G, Valverde A, Reynolds D et al (2015) Capacity Development. In: Worboys GL, Lockwood M, Kothari A, et al. (eds) Protected Area Governance and Management. ANU Press, Canberra, Australia, pp 251–291

Mullins RB, McKeown NJ, Sauer WHH, Shaw PW (2018) Genomic analysis reveals multiple mismatches between biological and management units in yellowfin tuna (*Thunnus albacares*). ICES J Mar Sci 75:2145–2152. https://doi.org/10.1093/icesjms/fsy102

Myosho T, Takehana Y, Hamaguchi S, Sakaizumi M (2015) Turnover of Sex Chromosomes in *Celebensis* Group Medaka Fishes. G3 Genes|Genomes|Genetics 5:2685–2691. https://doi.org/10.1534/g3.115.021543

Nakiyende H, Taabu-Munyaho A, Nansereko F, Mbabazi D (2018) Socio-economic and sustainability consequences of the decline of large fish species in the Ugandan large lakes. In: IIFET Conference. Oregon State University, Seattle, USA

Natri HM, Merilä J, Shikano T (2019) The evolution of sex determination associated with a chromosomal inversion. Nat Commun 10:145. https://doi.org/10.1038/s41467-018-08014-y

Nielsen E, Hemmer-Hansen J, Larsen P, Bekkevold D (2009) Population genomics of marine fishes : identifying adaptive variation in space and time. 18:3128–3150. https://doi.org/10.1111/j.1365-294X.2009.04272.x

Nielsen EE, Cariani A, Aoidh E Mac, Maes G, Milano I, Ogden R, et al (2012) Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nat Commun 3:851. https://doi.org/10.1038/ncomms1845

Nirchio M, Rossi AR, Foresti F, Oliveira C (2014) Chromosome evolution in fishes: a new challenging proposal from Neotropical species. Neotropical Ichthyology 12:761–770. https://doi.org/10.1590/1982-0224-20130008

Ntakimazi G (2006a) *Limnothrissa miodon*. The IUCN Red List of Threatened Species. Accessed 13 August 2024

Ntakimazi G (2006b) *Stolothrissa tanganicae*. The IUCN Red List of Threatened Species. Accessed 13 August 2024

Obiero K, Lawrence T, Ives J, Smith S, Njaya F, Kayanda R, et al (2020) Advancing Africa's Great Lakes research and academic potential: Answering the call for harmonized, long-term, collaborative networks and partnerships. J Great Lakes Res 46:1240–1250. https://doi.org/10.1016/j.jglr.2020.02.002

O'Connell M, Donnison A, Lynch K, Bennett R (2022) Assessing national-level provision of conservation capacity building: lessons learnt from a case study of Kenya. Oryx 56:753–759. https://doi.org/10.1017/s0030605322000345

Ogden R (2008) Fisheries forensics: The use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. Fish Fish 9:462–472. https://doi.org/10.1111/j.1467-2979.2008.00305.x

Ogutu-Ohwayo R, Hecky RE (1991) Introductions in Africa and Some of Their Implications. Can J Fish Aquat Sci 48:8–12. https://doi.org/https://doi.org/10.1139/f91-299

Olden JD, Poff NL, Bestgen KR (2006) Life-history strategies predict fish invasions and extirpations in the Colorado River Basin. Ecol Monogr 76:25–40. https://doi.org/10.1890/05-0330

Osio GC, Orio A, Millar CP (2015) Assessing the vulnerability of Mediterranean demersal stocks and predicting exploitation status of un-assessed stocks. Fish Res 171:110–121. https://doi.org/10.1016/j.fishres.2015.02.005

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick P (2010) Conservation genetics in transition to conservation genomics. Trends Genet 26:177–187. https://doi.org/10.1016/j.tig.2010.01.001

Ovenden JR, Leigh GM, Blower DC, Jones A, Moore A, Bustamante C, et al (2016) Can estimates of genetic effective population size contribute to fisheries stock assessments? J Fish Biol 89:2505–2518. https://doi.org/10.1111/jfb.13129

Ovenden JR, Moore A (2016) Working on the interface to achieve two-way information flow between policy makers and genetic scientists: a perspective. In: Genomics for improved fisheries management and conservation: have the promises been fulfilled? 7th World Fisheries Congress, Busan, Korea

Pala I, Naurin S, Stervander M, Hasselquist D, Bensch S, Hansson B (2012) Evidence of a neo-sex chromosome in birds. Heredity 108:264–272. https://doi.org/10.1038/hdy.2011.70

Pan Q, Kay T, Depincé A, Adolfi M, Schartl M, Guiguen Y, et al (2021) Evolution of master sex determiners: TGF-β signalling pathways at regulatory crossroads. Philos Trans R Soc Lond B Biol Sci 376:20200091. https://doi.org/10.1098/rstb.2020.0091

Paugy D (1986) Révision systématique des Alestes et Brycinus africains (Pisces, Characidae). Collection Etudes et Thèses Editions de l'ORSTOM, Paris 1–295

Pavan-Kumar A, Gireesh-Babu P, Jaiswar AK, Chaudhari A, Krishna G, Lakra W (2016) DNA barcoding of marine fishes: Prospects and challenges. In: DNA Barcoding in Marine Perspectives: Assessment and Conservation of Biodiversity. Springer International Publishing, pp 285–299

Payseur BA, Presgraves DC, Filatov DA (2018) Introduction: Sex chromosomes and speciation. In: Molecular Ecology. Blackwell Publishing Ltd, pp 3745–3748

Peck MA, Alheit J, Bertrand A, Catalán I, Garrido S, Moyano M, et al (2021) Small pelagic fish in the new millennium: A bottom-up view of global research effort. Prog Oceanogr 191:102494. https://doi.org/10.1016/j.pocean.2020.102494

Peichel CL, McCann SR, Ross JA, Naftaly A, Urton J, Cech J, et al (2020) Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome Biol 21:177. https://doi.org/10.1186/s13059-020-02097-x

Pereira LHG, Hanner R, Foresti F, Oliveira C (2013) Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? BMC Genet 14:20. https://doi.org/10.1186/1471-2156-14-20

Petrou EL, Fuentes-Pardo AP, Rogers LA, Orobko M, Tarpey C, Jiménez-Hidalgo I, et al (2021) Functional genetic diversity in an exploited marine species and its relevance to fisheries management. Proc R Soc B: Biol Sci 288:20202398. https://doi.org/10.1098/rspb.2020.2398

Pettersson M, Rochus CM, Han F, Chen J, Hill J, Wallerman O, et al (2019) A chromosome-level assembly of the Atlantic herring – detection of a supergene and other signals of selection. Genome Res 29:1–10. https://doi.org/10.1101/gr.253435.119

Pinsky ML, Palumbi SR (2014) Meta-analysis reveals lower genetic diversity in overfished populations. Mol Ecol 23:29–39. https://doi.org/10.1111/mec.12509

Plisnier PD, Nshombo M, Mgana H, Ntakimazi G (2018) Monitoring climate change and anthropogenic pressure at Lake Tanganyika. J Great Lakes Res. https://doi.org/10.1016/j.jglr.2018.05.019

Pomeroy R, Parks J, Courtney K, Mattich N (2016) Improving marine fisheries management in Southeast Asia: Results of a regional fisheries stakeholder analysis. Mar Policy 65:20–29. https://doi.org/10.1016/j.marpol.2015.12.002

Primmer CR (2009) From conservation genetics to conservation genomics. Ann N Y Acad Sci 1162:357–368. https://doi.org/10.1111/j.1749-6632.2009.04444.x

Qvarnström A, Bailey RI (2009) Speciation through evolution of sex-linked genes. Heredity 102:4–15. https://doi.org/10.1038/hdy.2008.93

Rafajlović M, Rambla J, Feder JL, Navarro A, Faria R (2021) Inversions and genomic differentiation after secondary contact: When drift contributes to maintenance, not loss, of differentiation. Evolution 75:1288–1303. https://doi.org/10.1111/evo.14223

Reichwald K, Petzold A, Koch P, Downie B, Hartmann N, Pietsch S, et al (2015) Insights into Sex Chromosome Evolution and Aging from the Genome of a Short-Lived Fish. Cell 163:1527–1538. https://doi.org/10.1016/j.cell.2015.10.071

Reiss H, Hoarau G, Dickey-Collas M, Wolff WJ (2009) Genetic population structure of marine fish: Mismatch between biological and fisheries management units. Fish Fish 10:361–395. https://doi.org/10.1111/j.1467-2979.2008.00324.x

Reynolds JR (1999) Fishcode management: Lake Tanganyika framework fisheries management plan. GCP/INT/648/NOR, Field Report F-2 (En). FAO. Rome, Italy

Rice WR (1984) Sex chromosomes and the evolution of sexual dimorphism. Evolution 38:735–742. https://doi.org/10.1111/j.1558-5646.1984.tb00346.x

Rick JA, Junker J, Kimirei IA, Sweke E, Mosille J, Dinkel C, et al (2022) The Genetic Population Structure of Lake Tanganyika's *Lates* Species Flock, an Endemic Radiation of Pelagic Top Predators. J Hered 113:145–159. https://doi.org/10.1093/jhered/esab072

Rius M, Turon X, Bernardi G, Volckaert F, Viard F (2015) Marine invasion genetics: from spatio-temporal patterns to evolutionary outcomes. Biol Invasions 17:869–885. https://doi.org/10.1007/s10530-014-0792-0

Robinson JPW, Mills DJ, Asiedu GA, Byrd K, Mancha Cisneros M, Cohen P, et al (2022) Small pelagic fish supply abundant and affordable micronutrients to low- and middle-income countries. Nat Food 3:1075–1084. https://doi.org/10.1038/s43016-022-00643-3

Rosa R, Caetano-Filho M, Shibatta OA, Giuliano-Caetano L (2009) Cytotaxonomy in distinct populations of *Hoplias* aff. *malabaricus* (Characiformes, Erythrinidae) from lower Paranapanema River basin. J Fish Biol 75:2682–2694. https://doi.org/10.1111/j.1095-8649.2009.02467.x

Ross JA, Urton JR, Boland J, Shapiro M, Peichel C (2009) Turnover of Sex Chromosomes in the Stickleback Fishes (Gasterosteidae). PLoS Genet 5:e1000391. https://doi.org/10.1371/journal.pgen.1000391

Rundle HD, Nosil P (2005) Ecological speciation. Ecol Lett 8:336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Ruzzante DE, Mariani S, Bekkevold D, André C, Mosegaard H, Clausen L, et al (2006) Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. Proc R Soc B: Biol Sci 273:1459–1464. https://doi.org/10.1098/rspb.2005.3463

Ryman N, Utter F, Laikre L (1995) Protection of intraspecific biodiversity of exploited fishes. Rev Fish Biol Fish 5:417–446. https://doi.org/https://doi.org/10.1007/BF01103814

Sako A, O'Reilly CM, Hannigan R, Bickford N, Johnson R (2005) Variations in otolith elemental compositions of two clupeid species, *Stolothrissa tanganicae* and *Limnothrissa miodon* in Lake Tanganyika. Geochem: Explor Environ Anal 5:91–97. https://doi.org/10.1144/1467-7873/03-039

Salzburger W, Van Bocxlaer B, Cohen AS (2014) Ecology and Evolution of the African Great Lakes and Their Faunas. Annu Rev Ecol Evol Syst 45:519–545. https://doi.org/10.1146/annurev-ecolsys-120213-091804

Sarropoulou E, Fernandes JMO (2011) Comparative genomics in teleost species: Knowledge transfer by linking the genomes of model and non-model fish species. Comp Biochem Physiol Part D Genomics Proteomics 6:92–102. https://doi.org/10.1016/j.cbd.2010.09.003

Satoh TP, Miya M, Mabuchi K, Nishida M (2016) Structure and variation of the mitochondrial genome of fishes. BMC Genomics 17:1–20. https://doi.org/10.1186/s12864-016-3054-y

Schindler DE, Armstrong JB, Reed TE (2015) The portfolio concept in ecology and evolution. Front Ecol Environ 13:257–263. https://doi.org/10.1890/140275

Scholz CA, Johnson TC, Cohen AS, King J, Peck J, Overpeck J, et al (2007) East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. PNAS 104:16416–16421. https://doi.org/10.1073/pnas.0703874104

Schwermer H, Barz F, Zablotski Y (2018) A Literature Review on Stakeholder Participation in Coastal and Marine Fisheries. In: Jungblut S, Liebich V, Bode-Dalby M (eds) YOUMARES 9 - The Oceans: Our Research, Our Future. Proceedings of the 2018 conference for YOUng MArine RESearcher, Oldenburg, Germany

Seehausen O, Wagner CE (2014) Speciation in Freshwater Fishes. Annu Rev Ecol Evol Syst 45:621–51. https://doi.org/10.1146/annurev-ecolsys-120213-091818

Shaklee JB, Beacham TD, Seeb L, White BA (1999) Managing fisheries using genetic data: case studies from four species of Pacific salmon. Fish Res 43:45–78. https://doi.org/10.1016/S0165-7836(99)00066-1

Sharpe DMT, Chapman LJ (2018) Contemporary phenotypic change in correlated characters in the African cyprinid, *Rastrineobola argentea*. Biol J Linn Soc 124:85–98. https://doi.org/10.1093/biolinnean/bly027

Sharpe DMT, Wandera SB, Chapman LJ (2012) Life history change in response to fishing and an introduced predator in the east African cyprinid *Rastrineobola argentea*. Evol Appl 5:677–693. https://doi.org/10.1111/j.1752-4571.2012.00245.x

Silva G, Cunha RL, Ramos A, Castilho R (2017) Wandering behaviour prevents inter and intra oceanic speciation in a coastal pelagic fish. Sci Rep 7:2893. https://doi.org/10.1038/s41598-017-02945-0

Simfukwe K, Limuwa MM, Njaya F (2022) Are Chilimira Fishers of *Engraulicypris sardella* (Günther, 1868) in Lake Malawi Productive? The Case of Nkhotakota District. Sustainability 14:16018. https://doi.org/10.3390/su142316018

Spliethoff PC, Longh HH, Frank VG (1983) Success of the introduction of the fresh water clupeid *Limnothrissa miodon* (Boulenger) in Lake Kivu. Aquac Res 14:17–31. https://doi.org/10.1111/j.1365-2109.1983.tb00050.x

Stephenson RL, Smedbol RK (2001) Small pelagic species fisheries. In: Steele JH (ed) Encyclopedia of Ocean Sciences, vol. 5, 1st edn. Elsevier Ltd, pp 468–473

Strømme C (1999) Studies on population structure of *Microthrissa moeruensis* (Poll, 1948), a small Clupeidae endemic to Lake Mweru, Zambia. [PhD thesis] University of Bergen, Norway

Symonová R, Majtánová Z, Sember A, Staaks G, Bohlen J, Freyhof J, et al (2013) Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. BMC Evol Biol 13:42. https://doi.org/10.1186/1471-2148-13-42

Takyi E (2019) Population genetic structure of *Sardinella aurita* and *Sardinella madurensis* in the Eastern Central Atlantic region (CECAF) in West Africa. [Master thesis] University of Rhode Island, USA

Tessier A, Richard A, Masilya P, Mudakikwa E, Muzana A, Guillard J (2020) Spatial and temporal variations of *Limnothrissa miodon* stocks and their stability in Lake Kivu. J Great Lakes Res 46:1650–1660. https://doi.org/10.1016/j.jglr.2020.09.009

Theissinger K, Fernandes C, Formenti G, Bista I, Berg PR, Bleidorn C, et al (2023). How genomics can help biodiversity conservation. Trends Genet 39:545–559. https://doi.org/10.1016/j.tig.2023.01.005

Tiercelin J, Mondeguer A (1991) The geology of the Tanganyika Trough. In: Coulter GW (ed) Lake Tanganyika and its life. Oxford University Press, pp 7–48

Tiercelin J, Lezzar K (2002) A 300 Million Years History of Rift Lakes in Central and East Africa: An Updated Broad Review. In: Odada EO, Olago DO (eds) The East African Great Lakes: Limnology, Palaeolimnology and Biodiversity. Advances in Global Change Research, vol 12. Springer, Dordrecht, pp 3–60

Triay C, Courcelle M, Caminade P, Bezault E, Baroiller J, Kocher T, et al (2022) Polymorphism of Sex Determination Amongst Wild Populations Suggests its Rapid Turnover Within the Nile Tilapia Species. Front Genet 13:820772. https://doi.org/10.3389/fgene.2022.820772

Turner JL (1982) Biological studies on the pelagic ecosystem of Lake Malawi, Fishery Expansion Project, Malawi, Tech. Report Fl:DP/MLWY75/019

Turner TL, Hahn MW, Nuzhdin S V. (2005) Genomic islands of speciation in *Anopheles gambiae*. PLoS Biol 3:1572–1578. https://doi.org/10.1371/journal.pbio.0030285

UNDP (1994) Pollution control and other measures to protect biodiversity of Lake Tanganyika. RAF/192/G32

Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. Trends Ecol Evol 24:110–117. https://doi.org/10.1016/j.tree.2008.09.011

Valenzuela-Quiñoez F (2016) How fisheries management can benefit from genomics? Brief Funct Genomics 15:352–357. https://doi.org/10.1093/bfgp/elw006

Van Doorn GS, Kirkpatrick M (2007) Turnover of sex chromosomes induced by sexual conflict. Nature 449:909–912. https://doi.org/10.1038/nature06178

Van Sickel WA, Kominz MA, Miller KG, Browning JV (2004) Late Cretaceous and Cenozoic sea-level estimates: Backstripping analysis of borehole data, onshore New Jersey. Basin Res 16:451–465. https://doi.org/10.1111/j.1365-2117.2004.00242.x

van Zwieten PAM, Aarnink BHM, Kapasa CK, Ngalande P (1996) The biology, fishery and trade of “chisense” from Mweru-Luapula: status of present knowledge. Department of Fisheries, Nchelenge, Zambia

Vandamme S, Raeymaekers JAM, Maes GE, et al (2021) Reconciling seascape genetics and fisheries science in three codistributed flatfishes. Evol Appl 14:536–552. https://doi.org/10.1111/eva.13139

Vanhove M, Tessens B, Schoelinck C, Jondelius U, Littlewood T, Artois T, et al (2013) Problematic barcoding in flatworms: A case-study on monogeneans and rhabdocoels (Platyhelminthes). ZooKeys 365:355–379. https://doi.org/10.3897/zookeys.365.5776

Venkatesh B (2003) Evolution and diversity of fish genomes. Curr Opin Genet Dev 13:588–592

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos Trans R Soc Lond B Biol Sci 367:451–460. https://doi.org/10.1098/rstb.2011.0260

Volff JN, Schartl M (2002) Sex determination and sex chromosome evolution in the medaka, *Oryzias latipes*, and the platyfish, *Xiphophorus maculatus*. Cytogenet Genome Res 99:170–177. https://doi.org/10.1159/000071590

Wambugu PW, Henry R (2022) Supporting in situ conservation of the genetic diversity of crop wild relatives using genomic technologies. Mol Ecol 31:2207–2222. https://doi.org/10.1111/mec.16402

Wang J, Santiago E, Caballero A (2016) Prediction and estimation of effective population size. Heredity 117:193–206. https://doi.org/10.1038/hdy.2016.43

Wangechi (2016) Genetic diversity, population structure and morphological characterization of the silver cyprinid *Rastrineobola argentea* (Pellegrin) in Port Victoria, Mbita and Nyanza Gulf of Lake Victoria (Kenya). [PhD thesis] Jomo Kenyatta University of Agriculture and Technology, Kenya

Wangechi, Muigai AWT, Otieno Ouma S (2015) The Impact of Evolution and Socio-economics of Commercially Exploited Fish Stock: A Review on *Rastrineobola argentea* in Lake Victoria. J Food Secur 3:82–86. https://doi.org/10.12691/jfs-3-3-3

Waples RS, Punt AE, Cope JM (2008) Integrating genetic data into management of marine resources: How can we do it better? Fish Fish 9:423–449. https://doi.org/10.1111/j.1467-2979.2008.00303.x

Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. J Fish Biol 74:329–356. https://doi.org/10.1111/j.1095-8649.2008.02080.x

Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. J Fish Biol 44:213–232. https://doi.org/10.1111/j.1095-8649.1994.tb01200.x

Wellenreuther M, Bernatchez L (2018) Eco-Evolutionary Genomics of Chromosomal Inversions. Trends Ecol Evol 33:427–440. https://doi.org/10.1016/j.tree.2018.04.002

White M (1973) Animal Cytology and Evolution. Cambridge University Press, UK

Whitehead PJP (1985) Clupeoid fishes of the world (Suborder Clupeoidei): An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies and wolf-herrings. FAO, Rome, Italy

Willi Y, Kristensen TN, Sgro CM, Weeks A, Ørsted M, Hoffmann A (2022) Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. PNAS 119:e2105076119. https://doi.org/10.1073/pnas.2105076119

Wilson AB, Teugels GG, Meyer A (2008) Marine incursion: The freshwater herring of Lake Tanganyika are the product of a marine invasion into West Africa. PLoS One 3:e1979. https://doi.org/10.1371/journal.pone.0001979

Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. Evolution 65:1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x

Ying Y, Chen Y, Lin L, Gao T (2011) Risks of ignoring fish population spatial structure in fisheries management. Can J Fish Aquat Sci 68:2101–2120. https://doi.org/10.1139/F2011-116

Zanzi A, Martinsohn JT (2017) FishTrace: a genetic catalogue of European fishes. Database 2017: bax075. https://doi.org/10.1093/database/bax075

Zardoya R, Meyer A (1996) Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol Biol Evol 13:933–942. https://doi.org/10.1093/oxfordjournals.molbev.a025661

Zhou Q, Wang J, Huang L, et al (2008) Neo-sex chromosomes in the black muntjac recapitulate incipient evolution of mammalian sex chromosomes. Genome Biol 9:R98. https://doi.org/10.1186/gb-2008-9-6-r98

Paper I

Leona J. M. Milec, Maarten P. M. Vanhove, Fidel Muterezi Bukinga, Els L. R. De Keyzer, Vercus Lumami Kapepula, Pascal Mulungula Masilya, N'Sibula Mulimbwa, Catherine E. Wagner and Joost A. M. Raeymaekers. Complete mitochondrial genomes and updated divergence time of the two freshwater clupeids endemic to Lake Tanganyika (Africa) suggest intralacustrine speciation (2022). *BMC Ecology & Evolution* 22(127), 1-7. https://doi.org/10.1186/s12862-022-02085-8

This is an open-access article, licensed under a Creative Commons Attribution 4.0 International License

Deed - Attribution 4.0 International - Creative Commons

BMC Ecology and Evolution

RESEARCH

Open Access

Complete mitochondrial genomes and updated divergence time of the two freshwater clupeids endemic to Lake Tanganyika (Africa) suggest intralacustrine speciation

Leona J. M. Milec[1,2*], Maarten P. M. Vanhove[2,3], Fidel Muterezi Bukinga[4], Els L. R. De Keyzer[3,5], Vercus Lumami Kapepula[4,6], Pascal Mulungula Masilya[4,7], N'Sibula Mulimbwa[4], Catherine E. Wagner[8] and Joost A. M. Raeymaekers[1]

Abstract

Background: The hydrogeological history of Lake Tanganyika paints a complex image of several colonization and adaptive radiation events. The initial basin was formed around 9–12 million years ago (MYA) from the predecessor of the Malagarasi–Congo River and only 5–6 MYA, its sub-basins fused to produce the clear, deep waters of today. Next to the well-known radiations of cichlid fishes, the lake also harbours a modest clade of only two clupeid species, *Stolothrissa tanganicae* and *Limnothrissa miodon*. They are members of Pellonulini, a tribe of clupeid fishes that mostly occur in freshwater and that colonized West and Central-Africa during a period of high sea levels during the Cenozoic. There is no consensus on the phylogenetic relationships between members of Pellonulini and the timing of the colonization of Lake Tanganyika by clupeids.

Results: We use short-read next generation sequencing of 10X Chromium libraries to sequence and assemble the full mitochondrial genomes of *S. tanganicae* and *L. miodon*. We then use Maximum likelihood and Bayesian inference to place them into the phylogeny of Pellonulini and other clupeiforms, taking advantage of all available full mitochondrial clupeiform genomes. We identify *Potamothrissa obtusirostris* as the closest living relative of the Tanganyika sardines and confirm paraphyly for *Microthrissa*. We estimate the divergence of the Tanganyika sardines around 3.64 MYA [95% CI: 0.99, 6.29], and from *P. obtusirostris* around 10.92 MYA [95% CI: 6.37–15.48].

Conclusions: These estimates imply that the ancestor of the Tanganyika sardines diverged from a riverine ancestor and entered the proto-lake Tanganyika around the time of its formation from the Malagarasi–Congo River, and diverged into the two extant species at the onset of deep clearwater conditions. Our results prompt a more thorough examination of the relationships within Pellonulini, and the new mitochondrial genomes provide an important resource for the future study of this tribe, e.g. as a reference for species identification, genetic diversity, and macroevolutionary studies.

Keywords: Great Lakes, Clupeiformes, Mitogenome, Time calibration, Phylogenetics

*Correspondence: milecleona@hotmail.com

[1] Faculty of Biosciences and Aquaculture, Nord University, Universitetsalléen 11, 8026 Bodø, Norway
Full list of author information is available at the end of the article

Background

Lake Tanganyika has experienced a turbulent geological history of lake level fluctuations, shifting shorelines and transient hydrological connections, paving the way for

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

a complex sequence of colonisations that gave rise to a diverse freshwater fauna with a high degree of endemism [1]. The lake was originally formed by lateral expansion of the western branch of the East African rift, crossing the predecessor of the Malagarasi–Congo River around 9–12 million years ago (MYA). Only 5–6 MYA its water levels rose high enough for the sub-basins and the swampy areas in between to fuse into the deep clearwater lake of today [2–4]. The lake has experienced large water level fluctuations since then [5, 6]. These events are reflected in the evolutionary history of the organisms inhabiting the lake. For example, the adaptive radiations of the Tanganyika cichlid tribes happened over several stages, with some ancestral species colonizing the lake in the early stages of its formation, while others diversified later when the historical sub-basin lakes fused [7, 8] or following the depression of the northernmost sub-basin around 7–8 MYA [9]. Yet other lineages were initially thought to have colonized the lake at an even later stage, and thus established themselves in an already present adaptive radiation [10, 11]. Recent work, however, suggests that the cichlid radiation unfolded completely within the temporal and spatial confines of Tanganyika [9, 12–14].

Next to this textbook example of adaptive radiation, Lake Tanganyika also harbours a small clade of two endemic clupeid species, *Stolothrissa tanganicae* and *Limnothrissa miodon*. These clupeids are members of the African clupeid tribe Pellonulini, one of the most diverse freshwater radiations of Clupeiformes with 22 species in 11 genera, most occurring either on the West coast of Africa (distribution from Senegal down to Congo/Angola), or in the Congo River system and its tributaries and lakes [15, 16]. The members of Pellonulini are thought to be derived from a group of sardine-like species whose ancestors originated from the Atlantic West coast of Africa during a period of high sea levels between 30 and 50 MYA [16–18]. The exact route this radiation took through the Congo Basin is unknown, and the relationships between pellonuline taxa remain inconsistent in published clupeid phylogenies [17, 19–21].

The Tanganyika sardines are the fully pelagic, planktivorous, endemic *S. tanganicae* and the semi-pelagic, more opportunistic *L. miodon*, which is originally endemic to Lake Tanganyika but has also established in other lakes in Central Africa after anthropogenic introductions. Both species are important fisheries targets in Lake Tanganyika and provide food and livelihood for millions of people [22, 23]. The colonization and subsequent speciation of the Tanganyika sardines has only been explicitly addressed once [17], and estimated as part of larger phylogenies twice more [19, 20]. In these studies, estimates of their divergence time are based on minimum one and maximum three mitochondrial genes, and show substantial variation, the youngest being at 3.91 MYA and the oldest at 8 MYA with a large credibility interval (CI). Lake Tanganyika was formed 9–12 MYA, with the northern and southern sub-basins forming at 7–8 MYA and 2–4 MYA, respectively. The fusion of the sub-basins and onset of clearwater conditions is estimated at 5–6 MYA. Keeping these estimates in mind, a divergence time of the two sardine species of 8–10 MYA would mean that the lineage leading to *S. tanganicae* and *L. miodon* started to undergo speciation soon after entering the not yet connected sub-basins of the proto-lake. An older divergence time would indicate riverine speciation and subsequent colonization of the proto-lake. In contrast, a more recent divergence time would agree with intralacustrine speciation i.e. after the sub-basins of the lake connected to form the deep rift lake we see today.

Robust phylogenies and estimates of divergence time between lineages are crucial to understanding the relationship between geological or hydrological events, speciation and realised biodiversity. Mitochondrial genes are routinely used for this purpose [24], but single-gene datasets have limited ability to recover true phylogenetic relationships, especially in more closely related species [13, 25, 26] and tend to overestimate divergence time [27]. Whole mitochondrial genomes can contain phylogenetic information that is lost when targeting a single gene, and have yielded higher resolution and better supported phylogenies in recent studies of fish [28, 29] and other vertebrates [27, 30, 31], especially when investigating recently diverged or taxonomically diverse taxa [27].

In this study, we use short-read next generation sequencing (NGS) to sequence and assemble the complete mitochondrial genomes of *S. tanganicae* and *L. miodon.* We then use the new sequences, together with all available full mitochondrial clupeiform genomes, to build the first phylogeny of members of Pellonulinae to include all mitochondrial protein-coding genes (PCGs), rRNA-genes and the D-loop (control region). We revisit the phylogenetic relationships within Pellonulinae and estimate the divergence time of the Lake Tanganyika sardines with improved resolution. We discuss the results in the light of the geological history of Lake Tanganyika.

Results

New mitochondrial genomes, diversity and divergence

The mitogenome assemblies of *S. tanganicae* and *L. miodon* were 16,737 bp and 16,739 bp long, respectively. We annotated all 13 PCGs, 22 transfer RNA (tRNA) genes, 2 rRNA genes (total = 37 genes) as well as the control region (D-loop) in both assemblies in the typical fish and vertebrate mitochondrial gene order [32] (Fig. 1). Gene order analysis in CREx confirmed that all included clupeiforms follow the same gene order, except *Ilisha*

Fig. 1 Mitochondrial genomes of *Stolothrissa tanganicae* (upper) and *Limnothrissa miodon* (lower). Inner circle shading indicates GC-content. Outer circle: black regions are protein coding genes, red regions are tRNA genes, beige regions are rRNA genes, and brown is the D-loop. Regions closer to the centre are located on the minus strand

elongata, where tRNA-Pro and tRNA-Thr appeared transposed.

Nucleotide diversity (π), calculated based on alignments of mitochondrial genes between 107 clupeiform and 6 non-clupeiform fish species, showed peaks in the beginning and end of 16S rDNA, as well as in the genes coding for ND1, ND2, ND3, ND5 and ATP synthase membrane subunit 6 (ATP6) (Fig. 2). The genes coding for COI, COII, COIII and CYTB, along with some regions of 12S and 16S rDNA, were relatively less diverse. The alignments for the D-loop, ND4L, ND4 and ND6 genes contained too many gaps to accurately calculate nucleotide diversity and were excluded from this analysis.

Analysis of pairwise genetic distance of PCGs revealed that *S. tanganicae* and *L. miodon* were among the 0.3% (PCGs) or 3% (non-coding regions) most similar species of Clupeiformes (17th or 206th most similar out of 5778 pairwise comparisons, respectively), and were the two most similar species of Pellonulini (1st out of 28 pairwise comparisons). When considering non-coding regions, the Tanganyika species pair was only the 22nd most similar, with *L. miodon* being more similar to most other pellonulines than to *S. tanganicae*. Among-group comparisons showed that *L. miodon* and *S. tanganicae* were almost equally differentiated from the remaining pellonulines when considering PCGs only (distance ± SE for *S. tanganicae* = 0.208 ± 0.006, *L. miodon* = 0.210 ± 0.006), but that *S. tanganicae* was more than twice as differentiated when considering non-coding regions only (distance ± SE for *S. tanganicae* = 0.221 ± 0.011, *L. miodon* = 0.106 ± 0.005).

Phylogenetic analysis

Maximum likelihood (ML) and Bayesian inference (Figs. 3, 4) placed *S. tanganicae* and *L. miodon* together with the other members of Pellonulini with high statistical support. Within Pellonulini, several genera appeared non-monophyletic. The position of *Microthrissa royauxi* was unresolved in these phylogenies, but a Bayesian analysis using only Dorosomatinae placed it paraphyletically with *M. congica, Pellonula* and *Odaxothrissa losera* with high confidence (Additional file 2: Fig. S2). The Lake Tanganyika sardines formed a well-supported clade nested within *Potamothrissa,* with *P. obtusirostris* more closely related to the Tanganyika sardines than to *P. acutirostris.* In phylogenies resulting from taxon-reduced datasets focusing on Dorosomatinae (Additional file 2), we did not observe any major topological changes compared to those based on the complete dataset. However, some deeper nodes within this subfamily were resolved, such as the placement of *Sardinella lemuru* with other species of *Sardinella* (Additional file 2: Figs. S1, S2). In

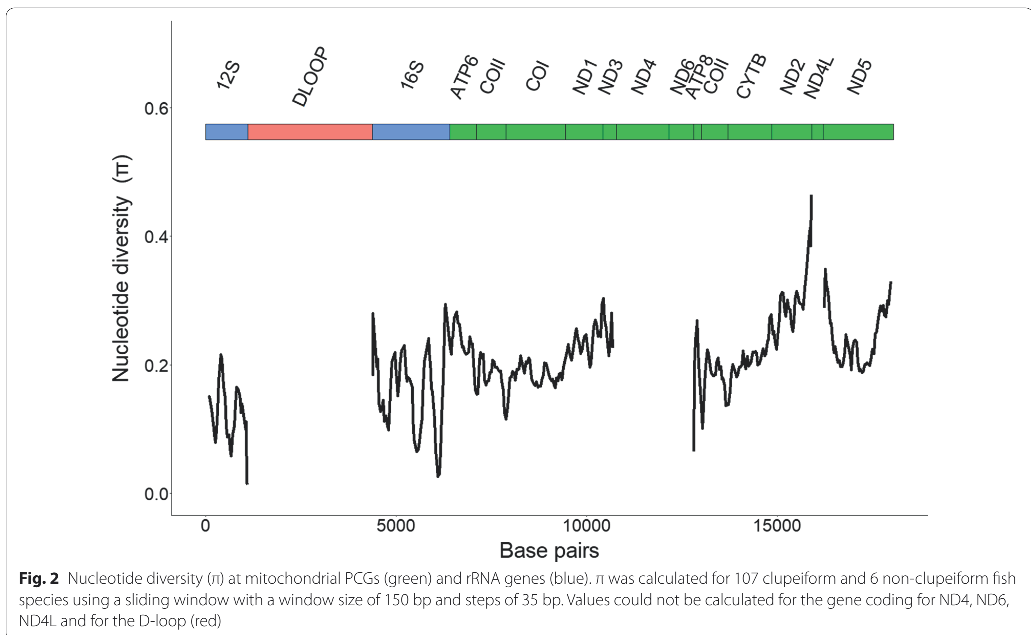

Fig. 2 Nucleotide diversity (π) at mitochondrial PCGs (green) and rRNA genes (blue). π was calculated for 107 clupeiform and 6 non-clupeiform fish species using a sliding window with a window size of 150 bp and steps of 35 bp. Values could not be calculated for the gene coding for ND4, ND6, ND4L and for the D-loop (red)

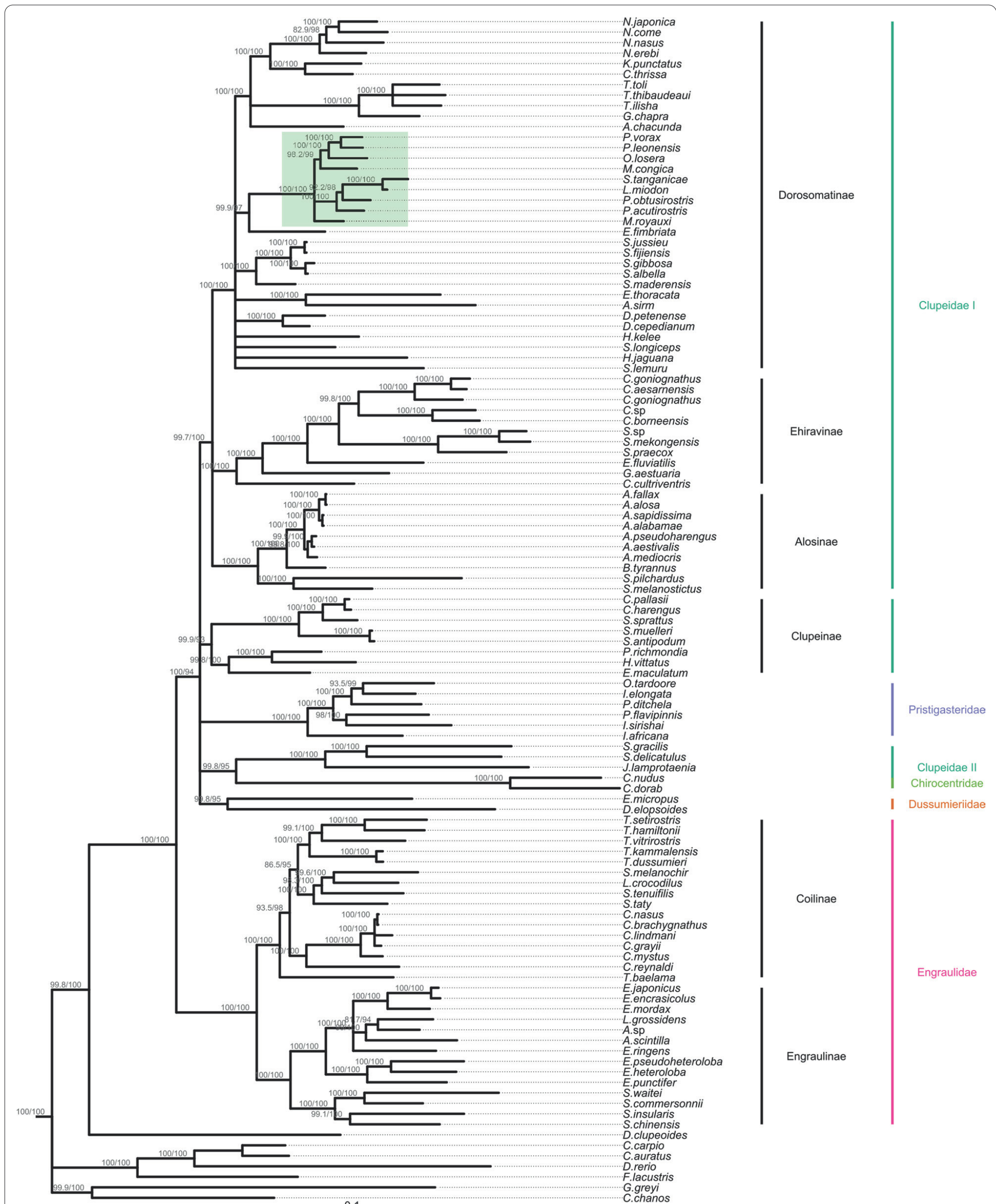

Fig. 3 Outgroup-rooted maximum likelihood phylogeny of Clupeiformes. Topology and branch lengths were estimated based on mitochondrial protein-coding genes, rRNA genes and D-loop sequence of 107 clupeiform and 6 non-clupeiform fishes. Node support was assessed by Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-aLRT%) and ultrafast bootstrap (UFBoot%). Nodes with SH-aLRT% < 75 and UFBoot% < 90 were polytomized and their support values are not shown. The scale bar indicates model-corrected evolutionary distance (expected number of nucleotide substitutions per site). Subfamilies are indicated on the right side in black, families in colour. Pellonulini is highlighted in green

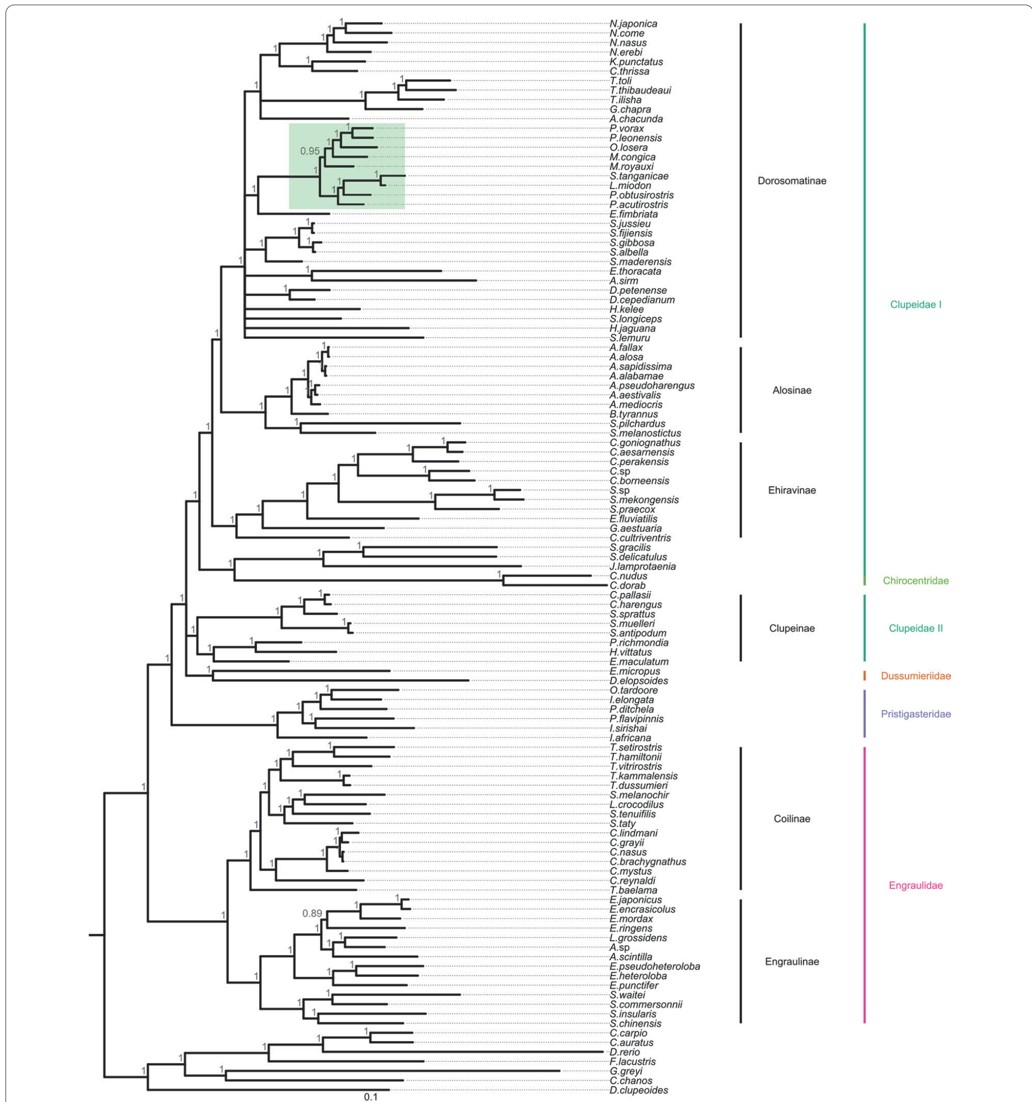

Fig. 4 Outgroup-rooted Bayesian phylogeny of Clupeiformes. Topology and branch lengths were estimated based on mitochondrial protein-coding genes, rRNA genes and D-loop sequence of 107 clupeiform and 6 non-clupeiform fishes. Node support was assessed by Bayesian posterior probabilities (BPP). Nodes with BPP < 0.85 were polytomized and their support values are not shown. Probabilities were rounded to the nearest 0.01. The scale bar indicates model-corrected evolutionary distance (expected number of nucleotide substitutions per site). Subfamilies are indicated on the right side in black, families in colour. Pellonulini is highlighted in green

addition, Bayesian inference including all Dorosomatinae and 21 other clupeiforms placed *Gudusia chapra* with species of *Tenualosa*, while *Anodontostoma chacunda* was placed with *Nematalosa, Konosirus punctatus* and *Clupanodon thrissa. Hilsa kelee, Dorosoma, Sardinella* and *Harengula jaguana* also clustered together in this

Bayesian phylogeny, and *Escualosa thoracata* and *Amblygaster sirm* formed a well-supported clade (Additional file 2: Fig. S2).

Outside of Dorosomatinae, all subfamilies of Clupeiformes, except Clupeinae, and most of their genera were retrieved with high support. Several deeper node placements in both trees, including some of the traditional clupeiform families with low taxonomic coverage, such as Pristigasteridae, Dussumieridae and Clupeidae II, had low support. Overall, ML and Bayesian analyses were in agreement, with the exception of those deeper, poorly supported nodes. We also found some genera split up or ambiguously placed. For example, there was a closer relationship between *Lycothrissa crocodilus* and *Setipinna melanochir* than the latter with other species of *Setipinna*. *Thryssa baleama* also did not cluster with other representatives of its own genus. *Ilisha elongata* clustered with *Pellona ditchela* and *Opisthopterus tardoore,* while *I. africana* and *I. sirishai* branched off earlier. Only two of the three species of *Sprattus* clustered together. The third, *S. sprattus,* was sister to the clade consisting of the two species of *Clupea*.

Dating of divergence time

Bayesian analysis estimated the divergence time of the Lake Tanganyika sardines at 3.64 MYA [95% CI: 0.99, 6.29] and the divergence between the most recent common ancestor (MRCA) of the Tanganyika sardines and their closest living relative in the tree, *P. obtusirostris,* at 10.92 MYA [95% CI: 6.37, 15.48]. The split between Pellonulini and the other clupeids and thus the timing of a large marine incursion into north-western Africa, was estimated at 43.71 MYA [95% CI: 31.79, 55.63] (Table 1, Fig. 5).

Discussion

We used NGS to sequence and assemble the complete mitochondrial genomes of the Tanganyika sardines, *S. tanganicae* and *L. miodon,* and built a phylogeny of Clupeiformes using full mitochondrial sequences with a focus on the West and Central-African tribe Pellonulini. Based on these complete mitogenomes, we estimated the divergence time of the Tanganyika sardines to investigate the timing of their speciation in relation to the geology of Lake Tanganyika.

Conserved gene order in Clupeiformes

Generally, mitochondrial gene arrangements have remained stable for long evolutionary times, but rearrangements do occur in many lineages of both invertebrates and vertebrates. Small rearrangements of neighbouring genes, for example clusters of tRNA-genes, and non-coding regions are especially common [33–35]. Several lineages of Actinopterygii are characterized by such rearrangements, but Clupeiformes is not one of them [32]. Our gene order analysis confirmed the conserved arrangement of mitochondrial genes in this order, aside from one transposition of two tRNA genes (tRNA-Pro and tRNA-Thr) in *Ilisha elongata.*

Inconsistent tree topologies within and outside Pellonulini

Both of our phylogenies (ML and Bayesian) support a single common ancestor for all included pellonuline species and recover *Ethmalosa fimbriata* as their sister species with high statistical support, consistent with Lavoué et al. [21]. This is in contrast with the results of Egan et al. [20] and Bloom and Lovejoy [19], neither of whom found good support for this sister-species relationship. *Microthrissa* appeared non-monophyletic in our study, and the Tanganyika sardines rendered *Potamothrissa* paraphyletic. The position of *Microthrissa* differs in almost every study that has addressed it. According to Egan et al. [20], *M. royauxi* is more closely related to the Tanganyika sardines and *Potamothrissa*, while *M. congica* clustered with *Pellonula* and *Odaxothrissa.* In the study of Wilson et al. [17], two specimens of *M. royauxi* did not even cluster together, but this may be a taxonomic artefact. Regardless of this possibility, *M. congica* was found more closely related to *Pellonula* than to *M. royauxi.* Bloom and Lovejoy [19], on the other hand, found both *Microthrissa* species more closely related to members of *Pellonula* and *Odaxothrissa* than to the clade including the Tanganyika sardines and *Potamothrissa.* In accordance, our analysis could not consistently recover the exact position of *M. royauxi* with high statistical support, but it did confirm that *M. congica* is more closely related to members of *Pellonula* and *Odaxothrissa* than to *M. royauxi.* The Lake Tanganyika sardines formed a well-supported clade nested within *Potamothrissa.* Contrarily, our study is the first to place *P. obtusirostris* and *P. acutirostris* in the same clade, albeit paraphyletically. With the improved resolution resulting from our whole mitogenome approach and the inclusion of a large number of clupeiform taxa, our study also confirms *E. fimbriata* as sister species of Pellonulini.

The source of these inconsistent topologies is unclear, but could be related to smaller, more variable and partly incomplete gene datasets in previous studies. Egan et al. [20] included only the gene coding for CYT-B for most members of Pellonulini, including the Tanganyika sardines, and three nuclear genes and/or the 16S rRNA gene for others. Bloom and Lovejoy [19] did not include *O. losera* and *P. acutirostris* and used CYT-B and 16S rRNA genes for most, and two nuclear genes for two species, while Wilson et al. [17] used only mitochondrial genes (CYT-B, 16S rRNA, 12S rRNA). None of the previous

Table 1 Comparison of key divergence times, taxa and markers in the pellonuline phylogeny between studies

	Divergence time estimate (MYA)				
	This study	Egan et al. 2018	Bloom and Lovejoy 2014	Wilson et al. 2008	Lavoué et al. 2013
Markers	**PCGs**	**CYT-B**, 16S, rag1, rag2, slc, zic1	**16S, CYT-B,** rag1, rag2	**16S, 12S, CYT-B**	PCGs, tRNAs, rRNAs
Number of taxa (excl. outgroup)	107	190	153	49	82
Number of sites (bp)	18,279	7135	5211	1049–1811	10,733
Node					
Limnothrissa miodon—Stolothrissa tanganicae	3.64 [0.99–6.29]	3.91 [1.19–6.64]	6.61 [2.20–11.01]	7.6 [2.1–15.9]	-
LT sardines– *Potamothrissa obtusirostris*	10.92 [6.37–15.48]	10.04 [5.62–14.47]	23.35 [16.37–30.33]	–	–
LT sardines– other pellonulines	–	–	–	27 [25.0–53.3]	–
Incursion 1: pellonulines–other clupeids	43.71 [31.79–55.63][1]	34.30 [25.56–43.03][1]	47.58 [35.68–59.47][1]	37 [25.0–53.3][2]	46.05 [33.38–58.71][1]
Incursion 2: *Gilchristella–Sauvagella*	–	25.00 [13.39–36.61]	33.92 [18.94–48.90]	20 [7.5–34.4]	–
Ehiravini–Pellonulini	64.57 [50.17–78.97]	70.13 [59.74–83.44]	98.24 [85.02–111.46]	48 [34.0–66.2]	89.02 [80.97–97.08]

Numbers between square brackets indicate 95% credibility intervals. Divergence times from our study were estimated in BEAST, those from other studies were directly reported or extracted from time-calibrated trees using WebPlotDigitizer. Markers indicated in bold were available for both *Stolothrissa tanganicae* and *Limnothrissa miodon*. PCGs = all mitochondrial protein coding genes, CYT-B = cytochrome B, 16S = 16S rRNA, 12S = 12S rRNA. [1]Split Pellonulini – Ethmalosa fimbriata. [2]Split Pellonulini—other clupeids (*E. fimbriata* not included in the study)

studies included more than around 5 kbp of alignment data, except Lavoué et al. [21], who included all PCG, rRNA and tRNA sequences which amounted to slightly over 10kbp. Taxonomic coverage was also highly variable, ranging from 49 to 190 clupeoid species, the lowest number belonging to Wilson et al. [17], which also had the largest credibility interval but had the highest taxonomic coverage of Pellonulini. The Tanganyika sardines and other pellonulines were also missing from several of these studies. Although the inclusion of taxa with incomplete datasets can help to resolve phylogenies [36], there is a trade-off with increased risk of phylogenetic artefacts, and difficulty detecting multiple substitutions [37, 38]. In contrast, our study had the first nearly complete dataset for all taxa, thanks to the readily available mitochondrial genomes from many pellonuline and other clupeid species.

Morphological diversity is relatively low in representatives of Pellonulini compared to for example cichlids [15, 39]. In the FAO species catalogue of clupeoid fishes, several ambiguous identifications and uncertain species descriptions are mentioned. For instance, the distinction between *O. losera* and *O. vittata* is based solely on the number of gill rakers, which also varies with the age of the specimen, a common occurrence among clupeid fishes. In *P. leonensis,* there is also evidence for undescribed subspecies exhibiting characteristics of both *P. leonensis* and *P. vorax,* or even specimens belonging to *Cynothrissa* [39]. It is thus not inconceivable that misidentifications have confounded past taxonomic studies, and that a taxonomic revision of these genera may be needed.

Outside of Pellonulini, we recovered most of the traditional families and subfamilies of Clupeiformes with high statistical support. The positions of the families with lower taxonomic coverage, including Pristigasteridae, Chirocentridae, Dussumieridae, remained unresolved, Clupeidae was not monophyletic and several species were also placed away from their congeners, for example in the genera *Setipinna*, *Thryssa, Ilisha* and *Sprattus.* These latter two findings are consistent with previous studies of Clupeiformes with taxonomic coverage comparable to ours [19, 20, 40], underlining the need for revision of these taxa.

Inconsistent divergence time estimates in Clupeiformes

Bayesian analysis estimated the divergence time of the Lake Tanganyika sardines at around 3.64 MYA [95% CI: 0.99, 6.29]. This estimate is younger than the previous estimate by Wilson et al. [17] at 7.6 MYA, but within its credibility interval [95% CI: 2.1, 15.9]. Conversely, their estimate fell just outside our credibility interval. Our estimate is also younger compared to the one by Bloom and Lovejoy [19] at 6.61 MYA [95% CI: 2.20, 11.01], but in accordance with Egan et al. [20] at 3.91 MYA [95% CI: 1.19, 6.64]. Our credibility intervals were smaller than

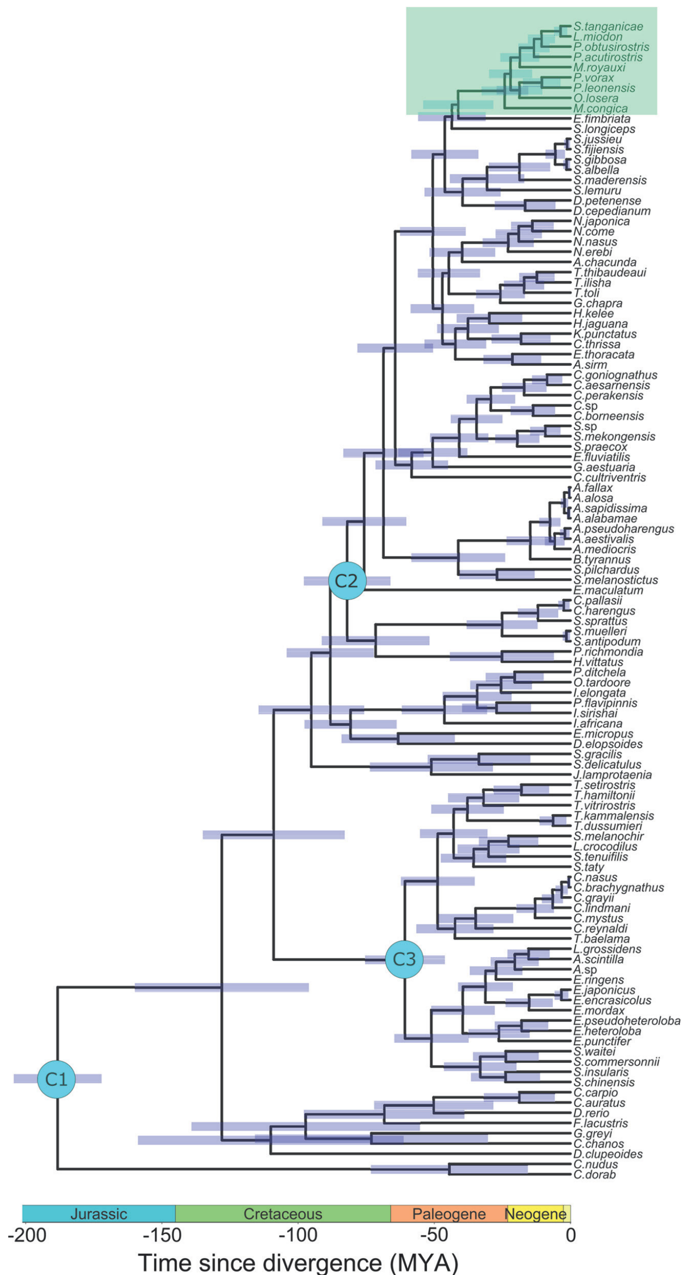

Fig. 5 Outgroup-rooted time-calibrated phylogeny of Clupeiformes. Divergence times were estimated using BEAST, based on the first and second codon positions of 13 mitochondrial protein coding genes of 107 clupeiform and 6 non-clupeiform fishes. Blue bars represent Bayesian 95% credibility intervals. Calibration points (C1–C3) are indicated on the corresponding nodes. Pellonulini is highlighted in green

those of Wilson et al. [17], but comparable to Bloom and Lovejoy [19] and Egan et al. [20]. Other, deeper nodes of interest also differed between the studies. Overall, our estimates most closely agreed with those of Egan et al. [20], except for the divergence time of the pellonulines from other clupeids, which was around 10 MYA older in our study. Bloom and Lovejoy [19] found consistently older estimates, while Wilson et al. [17] estimated the more recent nodes as older, and the deeper nodes as younger than the other three studies.

The differences in estimated divergence times between the studies can be partially attributed to the different estimation procedures. Specifically, methodological choices for Bayesian dating of nodes can strongly influence the accuracy and precision of the divergence time estimates, for example the choice of priors to account for uncertainty surrounding the age of a fossil, and the choice of clock model [31, 41]. Almost all the studies we compared here dated divergence using a fossil-calibrated uncorrelated (relaxed) clock model implemented in BEAST, accounting for substitution rate heterogeneity among branches. Six to eight fossil calibrations were specified as exponential priors with soft maximum ages. Only Wilson et al. [17] used an autocorrelated clock approach with seven fossil calibrations specified as uniform priors. In accordance with the three more recent studies on the divergence times of Clupeiformes [19–21], but in contrast with Wilson et al. [17], we chose a relaxed clock model, in accordance with the varying speeds of diversification in different clupeid lineages [40]. A possible caveat of our divergence time dating analysis is the exclusion of highly variable sequences of the D-loop region, rRNAs and third codon positions, which was necessary to achieve convergence of the model. These sites may be useful to resolve ambiguous recent divergences [42] and produce different, most likely slightly older, divergence time estimates.

Ideally, time-calibration of the diversification of the pellonulines would be based on fossils within this clade. Unfortunately, there are no known pellonuline fossils, and the fossil record of fishes of Central Africa in general is sparse [16]. In contrast to Wilson et al. [17], Bloom and Lovejoy [19] and Egan et al. [20], we decided to use pre-calibrated time scaling points ("secondary calibration") from a previously published study [43] instead of direct fossil calibration ("primary calibration"). Secondary calibrations can result in younger or older, depending on their placement, and falsely narrowed estimates of node ages [38, 44], but were necessary in our case to achieve convergence. Some caution is warranted when interpreting the width of our confidence intervals, but for the reasons described in the methods section 'Dating of divergence time', we are confident that our methodological choices have produced node age estimates with the best possible accuracy. The robustness of our approach is further supported by correspondence of our estimates to fossil ages from the literature. The ancestor of Clupeoidei was estimated at 127.48 [95% CI: 95.70, 159.27] in our study, older than the minimum age of 125 MYA attributed to the fossil of †*Cynoclupea nelsoni* [45]. The MRCA of Engraulidae was estimated at 60.10 MYA [95% CI: 44.70, 75.50], corresponding to the minimum age of the fossil of †*Eoengraulis fasoloi* of 50 MYA [46]. *Dorosoma petenense* was estimated as relatively old at 17.22 MYA [95% CI: 5.46, 28.97], but again within the boundaries of the minimum age of 2.5 MYA of the oldest fossil of *Dorosoma petenense* [47].

Utility of whole mitochondrial genomes for phylogenomic analysis and divergence time dating

Mitochondrial protein coding genes vary in their ability to recover known phylogenetic topologies. The sequences of ND4, ND5, COI and CYTB genes are generally useful for phylogenetic questions, while fast evolving genes such as ND4L and ATP8 are regarded as poor phylogenetic performers, although this differs per study and taxon [25, 26]. Indeed, we found relatively high nucleotide diversity in some genes or regions compared to others, including parts of the genes coding for the ATP6, ND1, ND2, ND3 and parts of ND5. However, whole mitochondrial genomes can recover accurate phylogenies with high resolution, despite containing "poor" phylogenetic performers [27, 48]. A smaller subset of "good" mitochondrial genes may be able to recover the same topology as the entire mitochondrial genome, but this is highly taxon-specific [25–27, 48]. Thus, utilizing more markers that provide complementary information is preferable if previous taxon-specific information on the utility of single markers is not available [27].

Phylogenies based on mitochondrial DNA (mtDNA) alone come with limitations [24, 49]. First, mtDNA is subject to frequent introgression, horizontal gene transfer, incomplete lineage sorting, and mitochondrial capture. As a result, past hybridization can go undetected in the absence of nuclear or morphological data [24, 50–54]. Second, sequencing or assembling nuclear pseudogenes of mitochondrial origin into mitogenomes can introduce false polyphyly within species or closely related taxa [55]. Third, the fast substitution rates of mtDNA make accurate estimation of deep divergences difficult due to problems with saturation and ensuing homoplasy [26, 27]. Overall, nuclear and mtDNA data can contrast or complement each other both in terms of tree topology and branch lengths [7, 8, 40]. Inconsistencies between them

should be considered informative, and future studies should strive to include both data sources to reconstruct more complete evolutionary scenarios [49].

There are several examples of ongoing hybridization between clupeid species [56–59]. With their similar habitat, nursery areas, and modes of reproduction, the Tanganyika sardines may well have a history of introgression that has remained concealed here. At the start of the Eocene (around 50 MYA), global sea levels were more than 100 m higher than today, and steadily decreased over the next 20 million years, with smaller maxima in between [16, 60, 61]. These fluctuations may have allowed frequent isolations and reconnections in the Congo Basin, favouring hybridization between other newly formed pellonuline species as well. To completely resolve the species tree of Pellonulini, phylogenomic analyses using nuclear genomic markers and multiple individuals per species are needed (but see Bloom & Egan [40], who found similar divergence time estimates with mtDNA and nuclear DNA datasets).

Updated divergence time suggests intralacustrine speciation of the Tanganyika sardines

Present-day distributions of several Afrotropical freshwater fish lineages show striking overlap, including members of Pellonulini, Kneriidae and Phractolaemidae, providing evidence for a single marine-freshwater transition across West- and Central Africa around 50 MYA during a period of high sea levels [16]. Despite the high sea levels, Lake Tanganyika was likely never in direct contact with the ocean and has not experienced much higher water levels than at present [3, 6]. Furthermore, due to uplift of the borders of the Congo Basin from the Cenozoic onwards, the possibility of an additional marine incursion close to the lake is faint [18]. It is therefore more likely the Lake Tanganyika sardines evolved from riverine clupeids. Indeed, the presence of a large body of water covering a large area of the Congo Basin ("paleolake Congo") until the Pliocene or early Pleistocene (2–12 MYA, [62, 63]), may have increased the connectivity between the Congo tributaries and its surrounding lakes, and may have facilitated the entry of riverine species into the predecessor of Lake Tanganyika at this time.

Our improved divergence time estimates of the Tanganyika sardines (3.64 MYA) and their MRCA from other pellonulines (10.92 MYA) help us to better understand their origin and colonisation time in connection to the geological history of the lake. Our estimates are compatible with (1) the entrance of the MRCA of the Tanganyika sardines into the newly formed Tanganyika basin (around 12 MYA) via the tributaries of the proto-Malagarasi-Congo River; and (2) intralacustrine speciation at the onset of deep- and clearwater conditions after the sub-basins fused (5–6 MYA). However, based on the 95% credibility intervals of our estimates, we cannot exclude the possibility that the MRCA of the Tanganyika sardines diverged from *P. obtusirostris* outside of the proto-lake and entered it sometime between the time of its formation and the fusion of its sub-basins.

Which environmental conditions triggered the divergence between *S. tanganicae* and *L. miodon* remains uncertain. Sexual selection, such as in cichlids [64], is unlikely to have played a large role due to the mode of reproduction of the clupeids. Ecological differences can be powerful drivers of speciation, even in (partial) sympatry [65, 66]. The newly fused basin, adding ecological heterogeneity to the ancestral sardine's environment, may have favoured dietary specialization through divergent selection on polymorphic trophic traits. Niche separation and divergence can then prompt genetic reproductive isolation if reinforced by spatial or temporal separation of spawning or lower hybrid fitness [65–67]. Indeed, contemporary populations of *L. miodon* seem to spawn all year round and mostly in the littoral, while populations of *S. tanganicae* exhibit clear peak spawning times in the pelagic [68]. This suggests that at some point during their divergence, spawning became more common in their respective preferred habitats. An alternative explanation is that their speciation was triggered by periods of allopatry [65, 67]. Given our credibility intervals and the frequent water-level fluctuations potentially separating and reconnecting the southern and central sub-basins of Lake Tanganyika several times, it is likely that ancestral sardine populations frequently occurred in partial or complete isolation.

A *Limnothrissa*-like ancestor of the Tanganyika sardines?

According to our ML and Bayesian phylogenies, the Tanganyika sardines are most closely related to *P. obtusirostris, P. acutirostris*, and *M. royauxi*, but not *M. congica*, the closest living relative being *P. obtusirostris*. Ecological studies of these species are sparse, hindering systematic comparison. Much of their distribution overlaps and, aside from *M. royauxi*, stretches across most of the Congo Basin all the way down to the Lukuga River, which was connected to the Malagarasi River east of Tanganyika around the time of the lake's formation. While *M. congica* and *P. acutirostris* occur in both rivers and lakes, *P. obtusirostris* and *M. royauxi* seem to be more strictly riverine [39]. The diet of *P. obtusirostris* and *M. congica* consists mostly of aquatic and terrestrial insects [39, 69], with occasional piscivory in *M. congica* [69]. In *M. congica*, strong seasonal effects of water level fluctuations on both diet and reproduction have been observed [69, 70]. Ecologically, *L. miodon*, with its generalist diet including

insects and small fishes, is more similar to the riverine pellonulines than *S. tanganicae*, which is a strict planktivore. In addition, individuals of *L. miodon* and species of *Potamothrissa* share a morphological feature that is otherwise rare in clupeid fishes: a row of saw-like teeth at the side of the lower jaw [39]. We thus suggest that the ancestral Tanganyika sardine shared more ecological traits with *L. miodon* than with *S. tanganicae.* This is also reflected in the more shorebound and generalist lifestyle of *L. miodon*, and its ability to invade the Cahora Bassa reservoir though dispersal via the riverine environment of the Zambezi [71], suggesting a relatively high ecological flexibility compared to *S. tanganicae* [72], and thus a higher ability to colonize a new environment. Nevertheless, the presence of established contemporary populations of *S. tanganicae* in one of the Congo's tributaries, the Lukuga, attest its ability to inhabit, or at least cross, non-pelagic environments, provided the water composition is sufficiently similar [73]. We also found larger genetic differentiation of *S. tanganicae* than *L. miodon* from the remaining pellonulines in non-coding regions. This could further support our hypothesis of a higher relatedness between *L. miodon* and the ancestral sardine, but may also indicate different demographic histories [74]. Kmentová et al. [75] found signatures of recent population expansion in both *L. miodon* and *S. tanganicae*, but these were more pronounced in the latter. The population expansion in *S. tanganicae* might be linked to the fusion of sub-basins, or any other major lake-level fluctuation that increased the amount of pelagic habitat. Similarly, species of the pelagic cichlid tribe Bathybatini showed recent demographic expansions, probably also linked to lake-level fluctuations [76].

Conclusion

Using NGS data, we assembled and annotated the full mitochondrial genomes of the Tanganyika sardines *S. tanganicae* and *L. miodon.* Putting them into phylogenetic context with full mitochondrial genomes of 107 other clupeiform species, we estimate their divergence time at 3.64 MYA, and divergence from their riverine ancestor at 10.92 MYA. This estimate implies that the MRCA of the Tanganyika sardines entered Lake Tanganyika shortly after its formation during a period of high connectivity of the Congo Basin's water bodies. We suggest that the speciation event is likely to have been brought on by the fusion of Lake Tanganyika's sub-basins and the subsequent clearwater conditions.

The mitochondrial genomes of *S. tanganicae* and *L. miodon* are valuable resources for future studies of the evolutionary history of these species at the population level, for example as a reference for barcoding, studies of their mitochondrial diversity and evolutionary history, as well as macroevolutionary study of relationships within Pellonulini and Clupeiformes. Future work should focus on the divergence time of different regions of the Tanganyika sardines' genomes and compare them to a dataset of nuclear genes or genome-wide data. This, in combination with formal tests for hybridization, could help to gauge the role of introgression in the timing and the scenario of speciation. Nuclear genomic sequences from several individuals of all members of Pellonulini would allow a more precise reconstruction of their colonization of West-Africa and clarify the ambiguous classifications in this group.

Methods

DNA extraction, library preparation and sequencing

One female individual of the two species was collected from liftnet fishing catches on the night of 15th of December 2018 off the shore of Uvira, Democratic Republic of Congo. Fish were dissected to extract liver tissue, which was directly frozen on dry ice and subsequently stored at − 20 °C until extraction. High molecular weight genomic DNA (gDNA) was extracted using a Blood and cell culture DNA Midi Kit (Qiagen). Libraries were prepared for each species separately using Chromium Genome Library & Gel Bead Kit v.2 (10X Genomics, cat. 120258), Chromium Genome Chip Kit v.2 (10X Genomics, cat. 120257), Chromium i7 Multiplex Kit (10X Genomics, cat. 120262) and Chromium controller according to the manufacturer's instructions with one modification (added shearing step before Illumina library preparation). Briefly, gDNA diluted to 1.02 ng/µl was combined with Master Mix, a library of Genome Gel Beads, and partitioning oil to create Gel Bead-in-Emulsions (GEMs) on a Chromium Genome Chip. The GEMs were isothermally amplified with primers containing an Illumina Read 1 sequencing primer, a unique 16 bp 10X barcode and a 6 bp random primer sequence. Barcoded DNA fragments were recovered for Illumina library construction. The amount and fragment size of post-GEM DNA was quantified using a Bioanalyzer 2100 with an Agilent High sensitivity DNA kit (Agilent, cat. 5067-4626). Prior to Illumina library construction, the GEM amplification product was sheared on an E220 Focused Ultrasonicator (Covaris, Woburn, MA) to approximately 350 bp (55 s at peak power = 175, duty factor = 10, and cycle/burst = 200). Then, the sheared GEMs were converted to a sequencing library following the 10X standard operating procedure. The library was quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems-Roche) and sequenced on a partial lane of the NovaSeq6000

sequencer (Illumina, San Diego, CA) with paired-end 150 bp reads.

Mitochondrial genome assembly

For mitogenome assembly, raw 10X Chromium reads were processed using the proc10xG package [77]. Process_10xReads.py was run using default settings to remove GEM and individual sample barcodes. The resulting reads passed read assessment by FastQC v0.11.7 [78] without any quality problems, residual adapters or over-represented sequences, the latter also commonly indicating adapter contamination. Mitogenomes of the two sardines were assembled from these barcode trimmed reads using MitoZ v.2.4-alpha [79] with default settings. Mitochondrial genes were annotated using the Mitofish annotator web service [80].

Taxonomic sampling and alignment

Taxonomic sampling for phylogenetic analysis included all members of Clupeiformes for which a complete mitochondrial genome is published (accessed 21st of October 2020, Additional file 1: Table S1). The sequences of *Odaxothrissa vittata* (NC_009590.1) and *Etrumeus teres* (NC_009583.1) were identical to *Pellonula vorax* and *E. micropus,* respectively, and were omitted from analysis. The outgroup was selected based on Lavoué et al. [21] and includes the denticle herring (*Denticeps clupeoides*), two alepocephaliforms, four ostariophysians and two euteleosts (Additional file 1: Table S1). We extracted mitogenomes and their annotations from NCBI using a combination of efetch [81] and custom Bash, Perl and Python scripts. We manually verified the new annotations of *S. tanganicae* and *L. miodon* by comparing the nucleotide and amino acid sequences, translated using vertebrate mitochondrial code, to the already published pellonuline mitogenomes, and checking for the presence of start and stop-codons at the appropriate positions in MEGA-11 [82].

We separately aligned each PCG using a codon-based MAFFT algorithm in the TranslatorX server [83]. We selected options for less stringent selection which allowed smaller final blocks with gap positions and less strict flanking positions. D-loop sequences were aligned using MAFFT v.7.470 with default parameters, and sequences of the rRNA coding regions using MAFFT with the -qinsi option [84]. Incomplete or missing regions were coded as missing data (N). We performed all alignments with and without alignment cleaning by Gblocks, further referred to as 'complete' and 'trimmed' alignments. We used AMAS v.0.98 [85] to separately concatenate the trimmed and complete alignments, producing one trimmed and one complete dataset. The phylogenetic content of these datasets was compared using likelihood mapping [86] implemented in TREE-PUZZLE v.5.3.rc16 [87]. We determined the optimal model for each dataset using jModelTest 2 [88] and specified these as input models for TREE-PUZZLE. Since there was no difference in phylogenetic content (86.6% fully resolved quartets, 2.3% partly resolved, 11.1% unresolved), we performed all subsequent analyses on the complete dataset of 18,279 bp.

Genetic diversity, divergence, and gene order

We calculated nucleotide diversity (π) of the final alignment using a sliding window analysis implemented in DnaSP v.6 [89] with a window size of 300 bp and steps of 15 bp. We used MEGA to quantify divergence between species and clades. We calculated pairwise genetic distances between all species, and mean between-group genetic distances between *S. tanganicae, L. miodon* and the remaining members of Pellonulini and Clupeiformes (in each of these comparisons excluding the other Tanganyika clupeid). The distances were calculated separately for PCGs (vertebrate mitochondrial code) and non-coding regions using a Tamura-Nei model including transitions and transversions, gamma-distributed rate variation among sites and heterogenous rate patterns among lineages. Gaps and missing data were deleted in a pairwise manner. The gamma parameter was estimated separately for the PCG and non-coding dataset using jModelTest 2. To estimate the relative similarity of *S. tanganicae* and *L. miodon* compared to similarities among other clupeids, we ranked all pairwise genetic distances of (1) Clupeiformes and (2) Pellonulini and calculated in which percentile the Tanganyika sardines fell using R v.4.0.4 [90]. Finally, we compared the gene order of all species included in our study using the CREx web application [91]. Two species, *Alosa fallax* and *Hilsa kelee,* were missing several markers (genes), and were thus excluded from this analysis.

Phylogenomic tree building

We ran IQ-TREE v.1.6.12 [92] twice on the complete dataset. The first run determined the best partition scheme (option -m MF + MERGE), allowing different models of molecular evolution in different genes/regions and, for PCGs, at the different codon positions [93]. The second run first implemented ModelFinder [94] to find the optimal model of evolution for each partition found by the previous round (options -spp and -m MFP), then constructed a ML tree, and finally assessed nodal support by 10,000 ultrafast bootstraps (generating support value UFBoot%) and 1,000 Shimodaira-Hasegawa-like approximate likelihood ratio test replicates (generating support value SH-aLRT%). A clade can be considered well supported if UFBoot% ≥ 95 (corresponding to a ~ 95% chance that the clade is true) and SH-aLRT% ≥ 80 [95, 96].

Using the IQ-TREE-derived partition and models, we constructed a Bayesian phylogeny in MrBayes v.3.2.7a [97], allowing estimation of the model parameters for each partition separately (unlinked character state frequencies, substitution rates of the GTR + I + G model). Two independent runs with 4 MCMC chains ran for 60 million generations, sampling every 500 generations and discarding the first 25% as burn in. The remaining samples were used to calculate Bayesian posterior probabilities (BPP) for each node in order to assess nodal support. A clade is considered well supported if BBP ≥ 0.9. The models converged, as indicated by the average standard deviation of split frequencies approaching zero, the absence of a trend in log likelihood of the runs, an Effective Sample Size (ESS) > 200, and the Potential Scale Reduction Factor approaching 1 [98].

Dating of divergence time

We estimated branch lengths and divergence times between *S. tanganicae* and *L. miodon* and five other nodes of interest (Table 1) using Bayesian relaxed molecular clock analysis implemented in BEAST v.2.6.3 [99] with the tree topology from MrBayes as a starting tree. We conducted four independent BEAST runs of 50 million generations. Convergence was ensured by checking if ESS was higher than 200 for all parameters using Tracer v.1.7.1 [100]. Trees were summarized and annotated using the TreeAnnotator module in BEAST. All final trees were visualized using ggtree v.3.3.0.901 in R [101].

For the time calibration of our tree, we first attempted a "primary calibration", which relies solely on fossils, using the complete dataset. However, when this model failed to converge, even after hundreds of millions of generations, we modified our analysis in two ways. First, instead of primary calibration, we chose to apply "secondary calibration", which uses estimates from an already existing phylogeny. We used three calibration points from a recent phylogeny of the teleosts using more than 30 fossils [43]. We included the MRCA of members of Clupeiformes and our outgroup (including *Danio rerio, Cyprinus carpio* and *Chanos chanos*) at 194 MYA (**C1**), MRCA of members of Clupeinae and Alosinae at 73 MYA (**C2**), and MRCA of Engraulidae at 61 MYA (**C3**). The calibration points were implemented as normal prior distributions for the node ages in BEAST. Secondary calibration inevitably incorporates geological and fossil uncertainty along with uncertainties associated with the primary dataset. They tend to push node age estimates into the more recent direction and falsely narrow the credibility intervals, especially when using a single old secondary calibration [38, 44], but see Powell et al. [102]. In our case, we presume the associated problems to be minimal for four reasons. (1) We used secondary calibrations for both old and younger nodes, which should diminish the tendency to estimate other nodes as younger [44]. (2) In Bayesian analysis implemented in BEAST, lognormally or exponentially distributed prior distributions are most commonly used for primary calibrations. These produce younger node age estimates and narrower credibility intervals than uniform (and presumably normally distributed) priors [44], which are more commonly used for secondary calibrations. (3) We validated the robustness of our secondary calibration by comparing the results of the BEAST dating to primary fossil calibration points from the literature, see discussion section 'Inconsistent divergence time estimates in Clupeiformes'. (4) The true uncertainty associated with secondary calibrations from Hughes et al. [43] based on more than 30 well-characterized fossils is likely smaller than that produced by using only three or four primary fossils with lognormal distributions with large variance.

Second, we omitted the hypervariable D-loop sequence, regions coding for rRNA, and the third codon positions, leaving a site-reduced dataset with only first and second codon positions of all PCGs. We also merged the separate partitions found by IQ-TREE into two partitions containing first and second codon positions. Since variable regions can be informative for estimating recent divergences, we repeated the analysis with two taxon-reduced datasets focusing on Dorosomatinae, in which we kept these variable regions (Additional file 2).

Finally, we compared our divergence time estimates for six nodes of interest with published estimates [17, 19–21] (Table 1). Estimates and 95% CI that were not directly reported in these publications were extracted from time-calibrated trees using WebPlotDigitizer v.4.5 [103].

Abbreviations

MYA: Million years ago; COI: Cytochrome *c* oxidase subunit I; CYTB: Cytochrome B; rRNA: Ribosomal RNA; ND: NADH-ubiquinone oxidoreductase; ND4L: NADH-ubiquinone oxidoreductase chain 4L; ATP6: ATP synthase membrane subunit 6; mtDNA: Mitochondrial DNA; gDNA: Genomic DNA; GEMs: Gel bead-in-emulsions; tRNA: Transfer RNA; NGS: Next generation sequencing; PCGs: Protein-coding genes; ML: Maximum likelihood; MRCA: Most recent common ancestor; CI: Credibility interval; BPP: Bayesian posterior probabilities; ESS: Effective sample size; CRH-U: Centre de Recherche en Hydrobiologie-Uvira.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12862-022-02085-8.

Additional file 1. Taxonomic information, accession numbers and references of mitochondrial genomes used for phylogenetic analyses.

Additional file 2. Phylogenetic analysis and divergence time dating of taxon-reduced datasets focusing on Dorosomatinae.

Acknowledgements
We thank Dovetail genomics for sequencing of the NGS data, Mathieu Tachon, Amalia Mailli, Armando J. Cruz-Laufer for their help with custom scripts, and Michael Matschiner for advice on divergence time dating.

Author contributions
LJMM, MPMV and JAMR conceptualized the study. LJMM collected and analysed the sequencing data and wrote the draft with input and interpretation by MPMV and JAMR. FMB, ELRDK, VLK, PMM and NM coordinated and executed the sampling of the Tanganyika sardines. All authors read and revised the draft and approved the final manuscript.

Funding
LJMM is supported by her Nord University PhD grant 224000-131. LJMM and JAMR are further supported by JAMR's start-up grant at Nord University, covering sequencing costs. The Special Research Fund of Hasselt University finances LJMM (BOF19DOC31) and MPMV (BOF20TT06). MPMV is further supported by Czech Science Foundation standard project GA19-13573S and Hasselt University Global Minds project GM2O18INITO7. ELRDK was funded through VLIR-UOS (VLADOC scholarship NDOC2016PR006 and South Initiative project CD2018SIN218A101) and BOF Seal of Excellence of University of Antwerp.

Availability of data and materials
The datasets supporting the conclusions of this article are available in the NCBI Sequence Read Archive [http://www.ncbi.nlm.nih.gov/bioproject/860551] and GenBank [Accessions: OP022425, OP021863] repositories. Custom scripts are available on Github: https://github.com/lmilec/mitoprep.

Declarations

Ethics approval and consent to participate
Collection of specimens used in the study complied with institutional, national, and international guidelines. Fieldwork in the Democratic Republic of Congo was carried out with the approval of the Centre de Recherche en Hydrobiologie – Uvira (CRH-U), which falls under the Congolese Ministry for science and technology ("Ministère National de la Recherche Scientifique et Technologie") under mission statement 002/MINRST/CRH-U/2018. Samples were exported with an export permit from the CRH-U. Samples were imported into the USA under import permit 70881B by the Department of the Interior U.S. Fish and Wildlife Service, Office of Law Enforcement to Dovetail Genomics. No animals were killed for this study, dead specimens were obtained from fishermen on Lake Tanganyika.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Faculty of Biosciences and Aquaculture, Nord University, Universitetsalléen 11, 8026 Bodø, Norway. [2]Centre for Environmental Sciences, Research Group Zoology: Biodiversity and Toxicology, Hasselt University, Agoralaan Gebouw D, 3590 Diepenbeek, Belgium. [3]Laboratory of Biodiversity and Evolutionary Genomics, Department of Biology, KU Leuven, Charles Déberiotstraat 32, 3000 Leuven, Belgium. [4]Centre de Recherche en Hydrobiologie-Uvira (CRH-Uvira), Uvira, Sud-Kivu, Democratic Republic of Congo. [5]Evolutionary Ecology Group (EVECO), Universiteit Antwerpen, Campus Drie Eiken, Universiteitsplein 1, 2610 Wilrijk, Belgium. [6]Université Catholique de Louvain, Place Sainte Barbe 2, 1348 Louvain-la-Neuve, Belgium. [7]Unité d'Enseignement et de Recherche en Hydrobiologie Appliquée (UERHA)-ISP/Bukavu, Bukavu, Sud-Kivu, Democratic Republic of Congo. [8]University of Wyoming, 1000 E University Ave, Laramie, WY 82071, USA.

Received: 20 July 2022 Accepted: 20 October 2022
Published online: 03 November 2022

References

1. Salzburger W, Van Bocxlaer B, Cohen AS. Ecology and evolution of the African Great Lakes and their faunas. Annu Rev Ecol Evol Syst. 2014;45(1):519–45.
2. Cohen AS, Soreghan MJ, Scholz CA. Estimating the age of formation of lakes: an example from Lake Tanganyika, East African Rift system. Geology. 1993;21(6):511–4.
3. Tiercelin J, Mondeguer A. The geology of the Tanganyika Trough. In: Coulter GW, editor. Lake Tanganyika and its life. Oxford: Oxford University Press; 1991. p. 7–48.
4. Tiercelin J-J, Lezzar K-E. A 300 million years history of rift lakes in central and east Africa: an updated broad review. In: Odada EO, Olago DO, editors. The East African great lakes: limnology, palaeolimnology and biodiversity advances in global change research, vol. 12. Springer: Dordrecht; 2002. p. 3–60.
5. Cohen AS, Stone JR, Beuning KRM, Park LE, Reinthal PN, Dettman D, et al. Ecological consequences of early Late Pleistocene megadroughts in tropical Africa. Proc Natl Acad Sci U S A. 2007;104(42):16422–7.
6. Cohen AS, Lezzar KE, Tiercelin AJJ, Soreghan M. New palaeogeographic and lake-level reconstructions of Lake Tanganyika: implications for tectonic, climatic and biological evolution in a rift lake. Basin Res. 1997;9(2):107–32.
7. Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C. Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. Syst Biol. 2002;51(1):113–35.
8. Sturmbauer C, Salzburger W, Duftner N, Schelly R, Koblmüller S. Evolutionary history of the Lake Tanganyika cichlid tribe Lamprologini (Teleostei: Perciformes) derived from mitochondrial and nuclear DNA data. Mol Phylogenet Evol. 2010;57(1):266–84.
9. Meyer BS, Matschiner M, Salzburger W. Disentangling Incomplete lineage sorting and introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes. Syst Biol. 2017;66(4):531–50.
10. Klett V, Meyer A. What, if anything, is a Tilapia?—mitochondrial ND2 phylogeny of tilapiines and the evolution of parental care systems in the African cichlid fishes. Mol Biol Evol. 2002;19(6):865–83.
11. Koch M, Koblmüller S, Sefc KM, Duftner N, Katongo C, Sturmbauer C. Evolutionary history of the endemic Lake Tanganyika cichlid fish *Tylochromis polylepis*: a recent intruder to a mature adaptive radiation. J Zool Syst Evol Res. 2007;45(1):64–71.
12. Ronco F, Matschiner M, Böhne A, Boila A, Büscher HH, El Taher A, et al. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. Nature. 2021;589(7840):76–81.
13. Meyer BS, Matschiner M, Salzburger W. A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach. Mol Phylogenet Evol. 2015;83:56–71.
14. Irisarri I, Singh P, Koblmüller S, Torres-Dowdall J, Henning F, Franchini P, et al. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. Nat Commun. 2018;9(1):3159.
15. Gourène G, Teugels GG. Synopsis de la classification et phylogénie des Pellonulinae de l'Afrique Occidentale et Centrale (Teleostei; Clupeidae). J Afr Zool. 1994;108(1):77–91.
16. Lavoué S. Origins of Afrotropical freshwater fishes. Zool J Linn Soc. 2020;188(2):345–411.
17. Wilson AB, Teugels GG, Meyer A. Marine incursion: The freshwater herring of Lake Tanganyika are the product of a marine invasion into West Africa. PLoS One. 2008;3(4):e1979.
18. Giresse P. Mesozoic-Cenozoic history of the Congo Basin. J Afr Earth Sci. 2005;43(1–3):301–15.
19. Bloom DD, Lovejoy NR. The evolutionary origins of diadromy inferred from a time-calibrated phylogeny for Clupeiformes (herring and allies). Proc R Soc B Biol Sci. 2014;281(1778):20132081.
20. Egan JP, Bloom DD, Kuo CH, Hammer MP, Tongnunui P, Iglésias SP, et al. Phylogenetic analysis of trophic niche evolution reveals a latitudinal herbivory gradient in Clupeoidei (herrings, anchovies, and allies). Mol Phylogenet Evol. 2018;124(March):151–61.

21. Lavoué S, Miya M, Musikasinthorn P, Chen WJ, Nishida M. Mitogenomic evidence for an Indo-West Pacific origin of the Clupeoidei (Teleostei: Clupeiformes). PLoS One. 2013;8(2):e56485.
22. Coulter GW. Biomass, production, and potential yield of the Lake Tanganyika pelagic fish community. Trans Am Fish Soc. 1981;110(3):325–35.
23. Mölsä H, Reynolds JE, Coenen EJ, Lindqvist OV. Fisheries research towards resource management on Lake Tanganyika. Hydrobiologia. 1999;407:1–24.
24. Ballard JWO, Whitlock MC. The incomplete natural history of mitochondria. Mol Ecol. 2004;13(4):729–44.
25. Cummings MP, Otto SP, Wakeley J. Sampling properties of DNA sequence data in phylogenetic analysis. Mol Biol Evol. 1995;12(5):814–22.
26. Zardoya R, Meyer A. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol Biol Evol. 1996;13(7):933–42.
27. Duchêne S, Archer FI, Vilstrup J, Caballero S, Morin PA. Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. PLoS One. 2011;6(11):e27138.
28. Lv Y, Li Y, Ruan Z, Bian C, You X, Yang J, et al. The complete mitochondrial genome of *Glyptothorax macromaculatus* provides a well-resolved molecular phylogeny of the Chinese sisorid catfishes. Genes. 2018;9(6):1–13.
29. Yamanoue Y, Miya M, Matsuura K, Katoh M, Sakai H, Nishida M. A new perspective on phylogeny and evolution of tetraodontiform fishes (Pisces: Acanthopterygii) based on whole mitochondrial genome sequences: basal ecological diversification? BMC Evol Biol. 2008;8(1):1–14.
30. Sullivan KAM, Platt RN, Bradley RD, Ray DA. Whole mitochondrial genomes provide increased resolution and indicate paraphyly in deer mice. BMC Zool. 2017;2(1):1–6.
31. Zhang P, Papenfuss TJ, Wake MH, Qu L, Wake DB. Phylogeny and biogeography of the family Salamandridae (Amphibia: Caudata) inferred from complete mitochondrial genomes. Mol Phylogenet Evol. 2008;49(2):586–97.
32. Satoh TP, Miya M, Mabuchi K, Nishida M. Structure and variation of the mitochondrial genome of fishes. BMC Genomics. 2016;17(1):1–20.
33. Boore JL. Animal mitochondrial genomes. Nucleic Acids Res. 1999;27(8):1767–80.
34. Dowton M, Castro LR, Austin AD. Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: the examination of genome "morphology." Invertebr Syst. 2002;16(3):345–56.
35. Satoh TP, Sato Y, Masuyama N, Miya M, Nishida M. Transfer RNA gene arrangement and codon usage in vertebrate mitochondrial genomes: a new insight into gene order conservation. BMC Genomics. 2010;11(1):479.
36. Wiens JJ. Missing data and the design of phylogenetic analyses. J Biomed Inform. 2006;39(1):34–42.
37. Roure B, Baurain D, Philippe H. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol Biol Evol. 2013;30(1):197–214.
38. Schenk JJ. Consequences of secondary calibrations on divergence time estimates. PLoS One. 2016;11(1):e0148228.
39. Whitehead PJP. Clupeoid fishes of the world (Suborder Clupeoidei): an annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies and wolf-herrings. Rome, Italy: FAO; 1985.
40. Bloom DD, Egan JP. Systematics of clupeiformes and testing for ecological limits on species richness in a trans-marine/freshwater clade. Neotrop Ichthyol. 2018;16(3):e180095.
41. Battistuzzi FU, Filipski A, Hedges SB, Kumar S. Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. Mol Biol Evol. 2010;27(6):1289–300.
42. Simmons MP, Zhang LIB, Webb CT, Reeves A. How can third codon positions outperform first and second codon positions in phylogenetic inference? An empirical example from the seed plants. Syst Biol. 2006;55(2):245–58.
43. Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, et al. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc Natl Acad Sci U S A. 2018;115(24):6249–54.
44. Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, et al. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). Syst Biol. 2012;61(2):289–313.
45. Malabarba MC, Di Dario F. A new predatory herring-like fish (Teleostei: Clupeiformes) from the early Cretaceous of Brazil, and implications for relationships in the Clupeoidei. Zool J Linn Soc. 2017;180(1):175–94.
46. Marramà G, Carnevale G. An Eocene anchovy from Monte Bolca, Italy: the earliest known record for the family Engraulidae. Geol Mag. 2016;153(1):84–94.
47. Miller RR. First fossil record (Plio-Pleistocene) of Threadfin Shad, *Dorosoma petenense*, from the Gatuña formation of Southeastern New Mexico. J Paleontel. 1982;56(2):423–5.
48. Williams SM, McDowell JR, Bennett M, Graves JE, Ovenden JR. Analysis of whole mitochondrial genome sequences increases phylogenetic resolution of istiophorid billfishes. Bull Mar Sci. 2018;94(1):73–84.
49. Rubinoff D, Holland BS. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst Biol. 2005;54(6):952–61.
50. Chan KMA, Levin SA. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution (N Y). 2005;59(4):720–9.
51. Funk DJ, Omland KE. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol Syst. 2003;34:397–423.
52. Toews DPL, Brelsford A. The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol. 2012;21(16):3907–30.
53. Ballard JWO, Rand DM. The population biology of mitochondrial DNA and its phylogenetic implications. Annu Rev Ecol Evol Syst. 2005;36:621–42.
54. Perea S, Vukić J, Šanda R, Doadrio I. Ancient mitochondrial capture as factor promoting mitonuclear discordance in freshwater fishes: a case study in the genus *Squalius* (Actinopterygii, Cyprinidae) in Greece. PLoS One. 2016;11(12):e0166292.
55. Funk DJ, Omland KE. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol Syst. 2014;2003(34):397–423.
56. Anderson JD, Karel WJ. Genetic evidence for asymmetric hybridization between menhadens (*Brevoortia* spp.) from peninsular Florida. J Fish Biol. 2007;71(SUPPL B):235–49.
57. Jolly MT, Maitland PS, Genner MJ. Genetic monitoring of two decades of hybridization between allis shad (*Alosa alosa*) and twaite shad (*Alosa fallax*). Conserv Genet. 2011;12(4):1087–100.
58. Hasselman DJ, Argo EE, McBride MC, Bentzen P, Schultz TF, Perez-Umphrey AA, et al. Human disturbance causes the formation of a hybrid swarm between two naturally sympatric fish species. Mol Ecol. 2014;23(5):1137–52.
59. Alexandrino P, Faria R, Linhares D, Castro F, Le Corre M, Sabatié R, et al. Interspecific differentiation and intraspecific substructure in two closely related clupeids with extensive hybridization, *Alosa alosa* and *Alosa fallax*. J Fish Biol. 2006;69(SUPPL B):242–59.
60. Van Sickel WA, Kominz MA, Miller KG, Browning JV. Late Cretaceous and Cenozoic sea-level estimates: backstripping analysis of borehole data, onshore New Jersey. Basin Res. 2004;16(4):451–65.
61. Haq BU, Hardenbohl J, Vail PR. Chronology of fluctuating sea levels since the Triassic (250 million years ago to present). Science. 1987;235(4):1156–67.
62. Beadle LC. The inland waters of Africa. London: Longman; 1974.
63. Peters C, O'Brien EM. Palaeo-lake Congo: implications for Africa's late Cenozoic climate—some unanswered questions. In: Palaeoecology of Africa and the Surrounding Islands, Vol 27; 2001.
64. Seehausen O, Terai Y, Magalhaes IS, Carleton KL, Mrosso HDJ, Miyagi R, et al. Speciation through sensory drive in cichlid fish. Nature. 2008;455(7213):620–6.
65. Rundle HD, Nosil P. Ecological speciation. Ecol Lett. 2005;8(3):336–52.
66. Seehausen O, Wagner CE. Speciation in freshwater fishes. Annu Rev Ecol Evol Syst. 2014;45:621–51.
67. Bernatchez L, Renaut S, Whiteley AR, Derome N, Jeukens J, Landry L, et al. On the origin of species: Insights from the ecological genomics of lake whitefish. Philos Trans R Soc B Biol Sci. 2010;365(1547):1783–800.

68. Mulimbwa NT, Milec LJM, Raeymaekers JAM, Sarvala J, Plisnier P, Marwa B, et al. Spatial and seasonal variation in reproductive indices of the clupeids *Limnothrissa miodon* and *Stolothrissa tanganicae* in the Congolese waters of northern Lake Tanganyika. Belgian J Zool. 2022;15:13–31.
69. Mokoma GM, Devos L, Cardona L. Alimentación de *Microthrissa congica* (Osteichthyes: Clupeidae) en la cuenca alta del río Congo (África). Rev Biol Trop. 1999;47(4):1081–6.
70. Mokoma M. Contribution à la connaissance de la Biologie de reproduction et du régime alimentaire de *Microthrissa congica* (Regan, 1917, Pisces, Clupeidae) du Bassin du Zaire. 1989.
71. Bernacsek GM, Lopes S. Cahora Bassa (Mozambique). CIFA Tech Pap. 1984;10:21–4.
72. Marshall BE. Why is *Limnothrissa* miodon such a successful introduced species and is there anywhere else we should put it? In: Pitcher TJ, Hart PJB, editors. The impact of species changes in African Lakes. 1995. pp. 527–45.
73. Kullander SO, Roberts TR. Out of Lake Tanganyika: endemic lake fishes inhabit rapids of the Lukuga River. Ichthyol Explor Freshw. 2011;22(4):355–76.
74. Tajima F, Nei M. Genetic drift and estimation of effective population size. Genetics. 1981;98:625–40.
75. Kmentová N, Koblmüller S, Van Steenberge M, Raeymaekers JAM, Artois T, De Keyzer ELR, et al. Weak population structure and recent demographic expansion of the monogenean parasite *Kapentagyrus* spp. infecting clupeid fishes of Lake Tanganyika, East Africa. Int J Parasitol. 2020;50(6–7):471–86.
76. Koblmüller S, Zangl L, Börger C, Daill D, Vanhove MPM, Sturmbauer C, et al. Only true pelagics mix: comparative phylogeography of deepwater bathybatine cichlids from Lake Tanganyika. Hydrobiologia. 2018;832(1):93–103.
77. Settle ML. Proc10xG. 2017. p. https://github.com/ucdavis-bioinformatics/proc10x.
78. Andrews S et al. FastQC: a quality control tool for high throughput sequence data. 2010. https://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/. 2010.
79. Meng G, Li Y, Yang C, Liu S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res. 2019;29:1–7.
80. Iwasaki W, Fukunaga T, Isagozawa R, Yamada K, Maeda Y, Satoh TP, et al. Mitofish and mitoannotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. Mol Biol Evol. 2013;30(11):2531–40.
81. Kans J. E-utilities on the Unix Command Line. In: Entrez programming utilities help. Bethesda (MD): National Center for Biotechnology Information (US); 2013.
82. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol. 2021;38(7):3022–7.
83. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 2010;38(SUPPL. 2):7–13.
84. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
85. Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016;2016(1):e1660.
86. Strimmer K, Von Haeseler A. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc Natl Acad Sci U S A. 1997;94(13):6815–9.
87. Schmidt HA, Strimmer K, Vingron M, Von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 2002;18(3):502–4.
88. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and high-performance computing Europe PMC Funders Group. Nat Methods. 2012;9(8):772.
89. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302.
90. R core team. R: a language and environment for statistical computing. Vienna, Austria; 2020. https://www.r-project.org/.
91. Bernt M, Merkle D, Ramsch K, Fritzsch G, Perseke M, Bernhard D, et al. CREx: inferring genomic rearrangements based on common intervals. Bioinformatics. 2007;23(21):2957–8.
92. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
93. Chernomor O, Von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 2016;65(6):997–1008.
94. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
95. Minh BQ, Nguyen MAT, Von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30(5):1188–95.
96. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307–21.
97. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19(12):1572–4.
98. Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. Stat Sci. 1992;7(4):457–72.
99. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4):1–28.
100. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018;67(5):901–4.
101. Yu G. Using ggtree to visualize data on tree-like structures. Curr Protoc Bioinf. 2020;69(1):1–18.
102. Powell CLE, Waskin S, Battistuzzi FU. Quantifying the error of secondary vs distant primary calibrations in a simulated environment. Front Genet. 2020;11(March):1–9.
103. Rohatgi A. WebPlotDigitizer. Pacifica, California, USA; 2021. https://automeris.io/WebPlotDigitizer.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

List of previously published theses for PhD in Aquaculture / PhD in Aquatic Biosciences / PhD in Biosciences, Nord University

No. 1 (2011)
PhD in Aquaculture
Chris André Johnsen
Flesh quality and growth of farmed Atlantic salmon (*Salmo salar* L.) in relation to feed, feeding, smolt type and season
ISBN: 978-82-93165-00-2

No. 2 (2012)
PhD in Aquaculture
Jareeporn Ruangsri
Characterization of antimicrobial peptides in Atlantic cod
ISBN: 978-82-93165-01-9

No. 3 (2012)
PhD in Aquaculture
Muhammad Naveed Yousaf
Characterization of the cardiac pacemaker and pathological responses to cardiac diseases in Atlantic salmon (*Salmo salar* L.)
ISBN: 978-82-93165-02-6

No. 4 (2012)
PhD in Aquaculture
Carlos Frederico Ceccon Lanes
Comparative Studies on the quality of eggs and larvae from broodstocks of farmed and wild Atlantic cod
ISBN: 978-82-93165-03-3

No. 5 (2012)
PhD in Aquaculture
Arvind Sundaram
Understanding the specificity of the innate immune response in teleosts: Characterisation and differential expression of teleost-specific Toll-like receptors and microRNAs
ISBN: 978-82-93165-04-0

No. 6 (2012)
PhD in Aquaculture
Teshome Tilahun Bizuayehu
Characterization of microRNA during early ontogeny and sexual development of Atlantic halibut (*Hippoglossus hippoglossus* L.)
ISBN: 978-82-93165-05-7

No. 7 (2013)
PhD in Aquaculture
Binoy Rajan
Proteomic characterization of Atlantic cod skin mucosa – Emphasis on innate immunity and lectins
ISBN: 978-82-93165-06-04

No. 8 (2013)
PhD in Aquaculture
Anusha Krishanthi Shyamali Dhanasiri
Transport related stress in zebrafish: physiological responses and bioremediation
ISBN: 978-82-93165-07-1

No. 9 (2013)
PhD in Aquaculture
Martin Haugmo Iversen
Stress and its impact on animal welfare during commercial production of Atlantic salmon (*Salmo salar* L.)
ISBN: 978-82-93165-08-8

No. 10 (2013)
PhD in Aquatic Biosciences
Alexander Jüterbock
Climate change impact on the seaweed *Fucus serratus*, a key foundational species on North Atlantic rocky shores
ISBN: 978-82-93165-09-5

No. 11 (2014)
PhD in Aquatic Biosciences
Amod Kulkarni
Responses in the gut of black tiger shrimp *Penaeus monodon* to oral vaccine candidates against white spot disease
ISBN: 978-82-93165-10-1

No. 12 (2014)
PhD in Aquatic Biosciences
Carlo C. Lazado
Molecular basis of daily rhythmicity in fast skeletal muscle of Atlantic cod *(Gadus morhua)*
ISBN: 978-82-93165-11-8

No. 13 (2014)
PhD in Aquaculture
Joanna Babiak
Induced masculinization of Atlantic halibut (*Hippoglossus hippoglossus* L.): towards the goal of all-female production
ISBN: 978-82-93165-12-5

No. 14 (2015)
PhD in Aquaculture
Cecilia Campos Vargas
Production of triploid Atlantic cod: A comparative study of muscle growth dynamics and gut morphology
ISBN: 978-82-93165-13-2

No. 15 (2015)
PhD in Aquatic Biosciences
Irina Smolina
Calanus in the North Atlantic: species identification, stress response, and population genetic structure
ISBN: 978-82-93165-14-9

No. 16 (2016)
PhD in Aquatic Biosciences
Lokesh Jeppinamogeru
Microbiota of Atlantic salmon (*Salmo salar L.*), during their early and adult life
ISBN: 978-82-93165-15-6

No. 17 (2017)
PhD in Aquatic Biosciences
Christopher Edward Presslauer
Comparative and functional analysis of microRNAs during zebrafish gonadal development
ISBN: 978-82-93165-16-3

No. 18 (2017)
PhD in Aquatic Biosciences
Marc Jürgen Silberberger
Spatial scales of benthic ecosystems in the sub-Arctic Lofoten-Vesterålen region
ISBN: 978-82-93165-17-0

No. 19 (2017)
PhD in Aquatic Biosciences
Marvin Choquet
Combining ecological and molecular approaches to redefine the baseline knowledge of the genus Calanus in the North Atlantic and the Arctic Oceans
ISBN: 978-82-93165-18-7

No. 20 (2017)
PhD in Aquatic Biosciences
Torvald B. Egeland
Reproduction in Arctic charr – timing and the need for speed
ISBN: 978-82-93165-19-4

No. 21 (2017)
PhD in Aquatic Biosciences
Marina Espinasse
Interannual variability in key zooplankton species in the North-East Atlantic: an analysis based on abundance and phenology
ISBN: 978-82-93165-20-0

No. 22 (2018)
PhD in Aquatic Biosciences
Kanchana Bandara
Diel and seasonal vertical migrations of high-latitude zooplankton: knowledge gaps and a high-resolution bridge
ISBN: 978-82-93165-21-7

No. 23 (2018)
PhD in Aquatic Biosciences
Deepti Manjari Patel
Characterization of skin immune and stress factors of lumpfish, *Cyclopterus lumpus*
ISBN: 978-82-93165-21-7

No. 24 (2018)
PhD in Aquatic Biosciences
Prabhugouda Siriyappagouder
The intestinal mycobiota of zebrafish – community profiling and exploration of the impact of yeast exposure early in life
ISBN: 978-82-93165-23-1

No. 25 (2018)
PhD in Aquatic Biosciences
Tor Erik Jørgensen
Molecular and evolutionary characterization of the Atlantic cod mitochondrial genome
ISBN: 978-82-93165-24-8

No. 26 (2018)
PhD in Aquatic Biosciences
Yangyang Gong
Microalgae as feed ingredients for Atlantic salmon
ISBN: 978-82-93165-25-5

No. 27 (2018)
PhD in Aquatic Biosciences
Ove Nicolaisen
Approaches to optimize marine larvae production
ISBN: 978-82-93165-26-2

No. 28 (2019)
PhD in Aquatic Biosciences
Qirui Zhang
The effect of embryonic incubation temperature on the immune response of larval and adult zebrafish (*Danio rerio*)
ISBN: 978-82-93165-27-9

No. 29 (2019)
PhD in Aquatic Biosciences
Andrea Bozman
The structuring effects of light on the deep-water scyphozoan *Periphylla periphylla*
ISBN: 978-82-93165-28-6

No. 30 (2019)
PhD in Aquatic Biosciences
Helene Rønquist Knutsen
Growth and development of juvenile spotted wolffish (*Anarhichas minor*) fed microalgae incorporated diets
ISBN: 978-82-93165-29-3

No. 31 (2019)
PhD in Aquatic Biosciences
Shruti Gupta
Feed additives elicit changes in the structure of the intestinal bacterial community of Atlantic salmon
ISBN: 978-82-93165-30-9

No. 32 (2019)
PhD in Aquatic Biosciences
Peter Simon Claus Schulze
Phototrophic microalgal cultivation in cold and light-limited environments
ISBN: 978-82-93165-31-6

No. 33 (2019)
PhD in Aquatic Biosciences
Maja Karoline Viddal Hatlebakk
New insights into Calanus glacialis and C. finmarchicus distribution, life histories and physiology in high-latitude seas
ISBN: 978-82-93165-32-3

No. 34 (2019)
PhD in Aquatic Biosciences
Arseny Dubin
Exploration of an anglerfish genome
ISBN: 978-82-93165-33-0

No. 35 (2020)
PhD in Aquatic Biosciences
Florence Chandima Perera Willora Arachchilage
The potential of plant ingredients in diets of juvenile lumpfish (*Cyclopterus lumpus*)
ISBN: 978-82-93165-35-4

No. 36 (2020)
PhD in Aquatic Biosciences
Ioannis Konstantinidis
DNA hydroxymethylation and improved growth of Nile tilapia (*Oreochromis niloticus*) during domestication
ISBN: 978-82-93165-36-1

No. 37 (2021)
PhD in Aquatic Biosciences
Youngjin Park
Transcriptomic and cellular studies on the intestine of Atlantic salmon
Discovering intestinal macrophages using omic tools
ISBN: 978-82-93165-34-7

No. 38 (2021)
PhD in Aquatic Biosciences
Purushothaman Kathiresan
Proteomics of early embryonic development of zebrafish (Danio rerio)
ISBN: 978-82-93165-37-8

No. 39 (2021)
PhD in Aquatic Biosciences
Valentin Kokarev
Macrobenthic communities of sub-Arctic deep fjords: composition, spatial patterns and community assembly
ISBN: 978-82-93165-38-5

No. 40 (2021)
PhD in Aquatic Biosciences
Aurélien Delaval
Population genomics of a critically endangered data-deficient elasmobranch, the blue skate *Dipturus batis*
ISBN: 978-82-93165-39-2

No. 41 (2021)
PhD in Aquatic Biosciences
Isabel Sofía Abihssira García
Environmental impact of microplastics in relation to Atlantic salmon farming
ISBN: 978-82-93165-40-8

No. 42 (2022)
PhD in Aquatic Biosciences
Yousri Abdelmutalab Ahmed Abdelhafiz
Insights into the bacterial communities of Nile tilapia – core members and intergenerational transfer
ISBN: 978-82-93165-41-5

No. 43 (2022)
PhD in Aquatic Biosciences
Fredrik Ribsskog Staven
Interaction studies on lumpfish exposed to Atlantic salmon: behavioural observations and the underlying physiological and neurobiological mechanisms
ISBN: 978-82-93165-42-2

No. 44 (2022)
PhD in Aquatic Biosciences
Solveig Lysfjord Sørensen
Influence of feed ingredients and additives on mucosal health with focus on the intestine of Atlantic salmon (*Salmo salar*)
ISBN: 978-82-93165-43-9

No. 45 (2022)
PhD in Aquatic Biosciences
Apollo Marco Dalonos Lizano
Examining challenges in species-level taxonomy among Calanus copepods in the Northern seas using genome and transcriptome data
ISBN: 978-82-93165-44-6

No. 46 (2022)
PhD in Aquatic Biosciences
Sowmya Ramachandran
Ribosomal RNA, ribose methylation, and box C/D snoRNAs during embryonic development of teleosts zebrafish (*Danio rerio*) and medaka (*Oryzias latipes*)
ISBN: 978-82-93165-45-3

No. 47 (2023)
PhD in Aquatic Biosciences
William John Hatchett
The brown algal genus Fucus: A unique insight into reproduction and the evolution of sex-biased genes
ISBN: 978-82-93165-46-0

No. 48 (2023)
PhD in Aquatic Biosciences
Ying Yen
Application of ensiled *Saccharina latissima* and *Alaria esculenta* as feed: ensilbility, digestibility and bioactivity
ISBN: 978-82-93165-47-7

No. 49 (2023)
PhD in Aquatic Biosciences
Deepak Pandey
Marine macroalgae as an alternative, environment-friendly, and bioactive feeding resource for animals
ISBN: 978-82-93165-48-4

No. 50 (2023)
PhD in Aquatic Biosciences
Nimalan Nadanasabesan
Dietary approaches to improve mucosal health of Atlantic salmon (*Salmo salar*)
ISBN: 978-82-93165-49-1

No. 51 (2023)
PhD in Aquatic Biosciences
Likith Reddy Pinninti
Biosystematics and evolutionary genomics of deep-sea fish (lumpsuckers, snailfishes, and sculpins) (Perciformes: Cottoidei)
ISBN: 978-82-93165-50-7

No. 52 (2023)
PhD in Biosciences
Mie Prik Arnberg
Directed endozoochory: a hitchhiker’s guide to successful sexual reproduction in clonal ericaceous plants
ISBN: 978-82-93165-51-4

No. 53 (2023)
PhD in Aquatic Biosciences
Adnan Hussain Gora
Insights from a zebrafish model to combat dyslipidemia using microbe-derived bioactive compounds
ISBN: 978-82-93165-52-1

No. 54 (2023)
PhD in Aquatic Biosciences
Saima Rehman
Diet-induced inflammation in zebrafish and its alleviation by functional oligo- and polysaccharides
ISBN: 978-82-93165-53-8

No. 55 (2023)
PhD in Biosciences
Clara Isabel Wagner
Exploring the nuclear genome of the spiny dogfish (*Squalus acanthias*)
ISBN: 978-82-93165-54-5

No. 56 (2023)
PhD in Aquatic Biosciences
Miiro Ilmari Virtanen
Internal tagging in Atlantic salmon (*Salmo salar L*.): A study of welfare, wounds, and stress
ISBN: 978-82-93165-55-2

No. 57 (2023)
Md Golam Rbbani
PhD in Biosciences
Characterization of Circular RNAs in Nile Tilapia Muscle and Thermal Modulation of their Expression in Relation to Growth
ISBN: 978-82-93165-56-9

No. 58 (2023)
Èric Jordà Molina
PhD in Aquatic Biosciences
Spatio-temporal dynamics of soft-bottom macrobenthic communities in a rapidly changing Arctic: a case study of the Northwestern Barents Sea
ISBN: 978-82-93165-57-6

No. 59 (2024)
Shubham Varshney
PhD in Aquatic Biosciences
From nanoplastics to chemical pollutants: Exploring mixture toxicity in zebrafish
ISBN: 978-82-93165-58-3

No. 60 (2024)
PhD in Aquatic Biosciences
Charles Patrick Lavin
The cumulative impacts of climate change and fishing on marine communities
ISBN: 978-82-93165-59-0

No. 61 (2024)
PhD in Biosciences
Dhurba Adhikari
Evolutionary genomics of pelvic spine reduction in *Gasterosteus aculeatus*
ISBN: 978-82-93165-60-6

No. 62 (2024)
PhD in Biosciences
Betty Marine Normann Furulund
Molecular evolution of catalytic RNA introns in nuclear ribosomal DNA from eukaryotic microorganisms
ISBN: 978-82-93165-61-3

PhD in Biosciences no. 63 (2024)
Francesc Gordo Vilaseca
The effect of climate warming on the north-east Atlantic marine fish community
ISBN: 978-82-93165-62-0

PhD in Biosciences no. 64 (2024)
Joshua Lustracion Superio
Development of immune-reproductive indicators to enhance breeding functionality in domesticated spotted wolffish (Anarhichas minor)
ISBN: 978-82-93165-63-7

PhD in Biosciences no. 65 (2024)
Exploration of the lipidome of cold-adapted microalgae under abiotic stress.
Hirono Suzuki
ISBN: 978-82-93165-64-4