Diving into Diversity: The Complex Evolutionary History and Species Richness of the 'sawfin barbs' from Lake Edward and Adjacent Systems

Peer-reviewed author version

Maetens, Heleen; Decru, Eva; Boom, Arthur Francois; VRANKEN, Nathan; VAN STEENBERGE, Maarten & Snoeks, Jos (2024) Diving into Diversity: The Complex Evolutionary History and Species Richness of the 'sawfin barbs' from Lake Edward and Adjacent Systems. In: Evolutionary biology, 51 (3-4), p. 366–383.

DOI: 10.1007/s11692-024-09640-9
Handle: http://hdl.handle.net/1942/44426

Diving into diversity: the complex evolutionary history and species richness of the 'sawfin barbs' from Lake Edward and adjacent systems

Maetens Heleen[a,b*] (0000-0002-1623-9492), Decru Eva[c] (0000-0002-6969-6726), Boom François Arthur[a] (0000-0001-5724-2754), Vranken Nathan[a,b,d,e] (0000-0003-0728-4715), Van Steenberge Maarten[d,e] (0000-0002-6964-9014), & Snoeks Jos[a,b] (0000-0002-3258-8369)

[a]Section Vertebrates, Biology Department, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080, Tervuren, Belgium.

[b]Laboratory of Fish Diversity and Conservation, Department of Biology, University of Leuven, Ch. Deberiotstraat 32, 3000, Leuven, Belgium.

[c]Department of Biology, University of Leuven, Ch. Deberiotstraat 32, 3000, Leuven, Belgium.

[d]Operational Directorate Taxonomy and Phylogeny, Royal Belgian Institute of Natural Sciences, Vautierstraat 29, 1000, Brussels, Belgium.

[e]Research Group Zoology: Biodiversity and Toxicology, Centre for Environmental Sciences, Hasselt University, Agoralaan Building D, 3590, Diepenbeek, Belgium.

*Corresponding author at Laboratory of Biodiversity and Evolutionary Genomics, Department of Biology, University of Leuven, Ch. Deberiotstraat 32, 3000, Leuven, Belgium. E-mail adress: heleen.maetens@kuleuven.be

Abstract

Enteromius Cope, 1867 is a species-rich genus of small cyprinids endemic to Africa, which includes the 'sawfin barbs'. This study explored the species diversity of this group within the Lake Edward system, including adjacent areas that belong to the Lakes Albert and Victoria systems. We used a multifaceted approach encompassing mitochondrial and nuclear DNA analyses, including a molecular clock analysis, and morphometrics. Additionally, broader regional relationships were investigated by including 'sawfin barbs' from other parts of the East Coast ichthyofaunal province and the Nile Basin, and from the Congo Basin, into the molecular analyses. In contrast to the previously reported three species from the Lake Edward system and adjacent areas, the results showed a fourfold increase in the number of species, thereby indicating that the three species actually constituted species complexes. Within these complexes, a consistent geographic pattern unfolded: if one species occurred at higher altitudes of the Lake Edward system, another closely related species occupied lower altitudes near Lakes Edward and George. This geographic consistency suggested an allopatric mode of speciation. Intriguingly, the revealed Pliocene-Pleistocene origin of nearly all species of 'sawfin barbs' from the Lake Edward system and neighbouring regions largely predated the important geological events that reshaped the hydrology in the western rift. This study offers a more detailed insight into the evolutionary patterns of the African small barbs representing a very high and unrecognized species diversity, accompanied by little morphological but high genetic divergence between species, indicating intriguingly old species origins.

Keywords

East Africa – *Enteromius* – molecular clock – biogeography – speciation

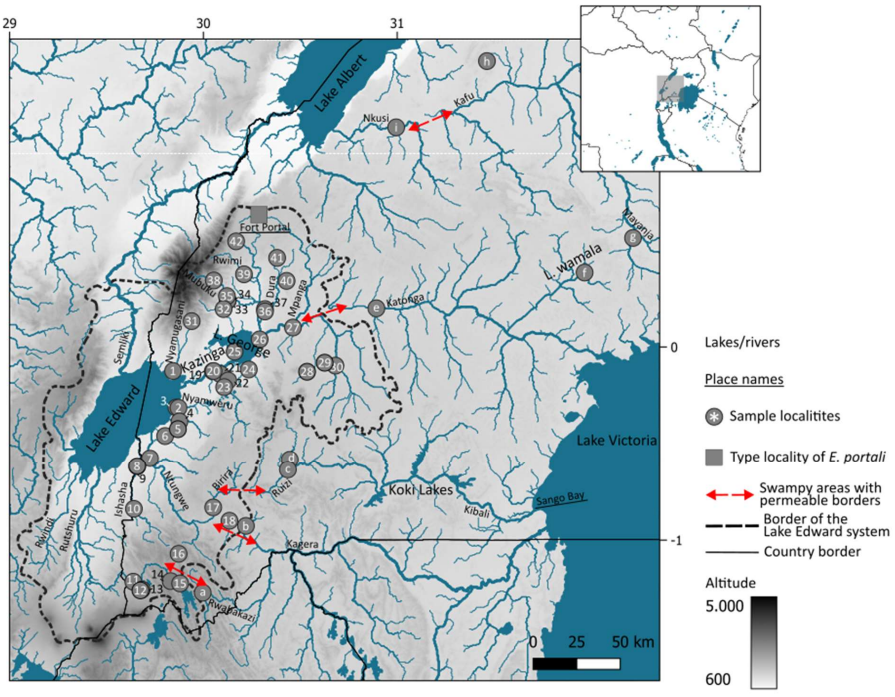
1. Introduction

Enteromius Cope, 1867, encompassing the diploid small-sized African barbs, forms one of the largest fish genera in the world with an estimated number of 218 valid species (Froese & Pauly, 2024). In Africa, it is only outnumbered by *Haplochromis* Hilgendorf, 1888, which contains 250 valid species (Froese & Pauly, 2024). The species diversity in *Enteromius* is expected to be much higher than reported. Species of *Enteromius* are often morphologically very similar and difficult to distinguish (e.g. Van Ginneken et al., 2017; Maetens et al., 2020). However, within the small African barbs, a distinction into three groups can be made based on the morphology of the third unbranched dorsal fin ray (Skelton et al., 1991; Maetens et al, 2020). The three morphological groups are here conveniently called 'sawfin barbs', 'spinefin barbs' and 'soft-rayed barbs', representing small barbs with an ossified dorsal spine with serrations, an ossified dorsal spine without serrations, and a smooth flexible dorsal spine, respectively. In spite of the practical use of this classification for identification purposes, recent phylogenetic reconstructions of *Enteromius* (e.g. Yang et al., 2015; Ren & Mayden, 2016; Hayes & Armbruster, 2017) did not support this classification as none of the three groups appeared to be monophyletic. Additionally, the genus *Enteromius* as currently defined is rendered paraphyletic by *Caecobarbus* Boulenger, 1921, *Clypeobarbus* Fowler, 1936, *Prolabeops* Schultz, 1941 and *Barboides* Brüning, 1929 (Yang et al., 2015; Ren & Mayden, 2016) and possibly also by *Barbopsis* di Caporiacco, 1926 and *Prolabeo* Norman, 1932, the phylogenetic placement of which is still pending because of a lack of genetic material (Tan & Armbruster, 2018; Schedel et al., 2022).

Recent studies (e.g. Schmidt et al., 2017, 2019; Van Ginneken et al., 2017; Englmaier et al., 2020; Decru et al., 2022) revealed that several *a priori* identified species, actually consisted of complexes of morphologically highly similar, but genetically distinct biological species. A striking example is the study by Van Ginneken et al. (2017), who found 23 mitochondrial lineages (cytochrome *c* oxidase subunit I, COI) within four *a priori* identified species from parts of the Congo Basin. Similar results were found in East African species: Schmidt et al. (2017) found unrecognised diversity within *E. kerstenii* (Peters, 1868) and *E. paludinosus* (Peters, 1852) based on analyses of the mitochondrial gene Cyt b (cytochrome b) and the nuclear genes RAG1 (recombination activating gene 1, exon 3) and intron 2 of GH (Growth Hormone). Englmaier et al. (2020) described a new species and revalidated another species within the *E. paludinosus* complex from the main Ethiopian Rift area. Also in West Africa, multiple candidate species were found within three distinct populations in *E. foutensis* (Lévêque, Teugels & Thys van den Audenaerde, 1988), based on Cyt b and RAG1 (Schmidt et al., 2019).

In this study, we investigated the 'sawfin' species of *Enteromius* from the Lake Edward system as defined by Decru et al. (2019). This system, which is situated in East Africa and which straddles the

77 border between the Democratic Republic of the Congo (DRC) and Uganda, consists of the basins of
78 Lakes Edward and George that are connected by the Kazinga Channel (Fig. 1). Lake Edward drains, via
79 the Semliki River and Lake Albert, to the Nile. Its drainage system includes several rivers, swamps and
80 isolated crater lakes. The Lake Edward system was initially assigned to the Nilo-Sudan Ichthyofaunal
81 Province by Roberts (1975), but later to the East Coast Ichthyofaunal Province (Greenwood, 1983;
82 Snoeks et al., 1997; Decru et al., 2019). The Lake Edward system has a complex hydrology that is
83 reminiscent of the region's turbulent past. The rapids on the Semliki River form a barrier with the
84 system of Lake Albert in the north (Beadle, 1974; Stewart, 2009; Decru et al., 2019). Mountain ranges
85 isolate the fish fauna at the system's western and southern boundaries, while a large swampy region
86 to the north-east of Lake George forms a permeable border with the system of Lake Victoria
87 (Doornkamp & Temple, 1966; Greenwood, 1973; Reardon & Chapman, 2009). This swampy region
88 includes the catchment of the Katonga River (Fig. 1). This river formed a connection during the
89 Pleistocene between Lake Victoria and Lake George. However, due to uplift of the western rift, the
90 major part of the Katonga River now drains into Lake Victoria, though a significant section of its flow is
91 still directed towards Lake George (Beadle, 1974; Doornkamp & Temple, 1966; Reardon & Chapman,
92 2009; Schraml & Tichy, 2010). More swampy areas that are dominated by papyrus can be found in the
93 southern part of the Lake Edward system where the direction of the flow, however, cannot be
94 determined (Fig. 1) (Beadle 1974). Swampy areas also straddle the border between the systems of
95 Lakes Albert and Victoria. To the east of Lake Albert, the Nkusi and Kafu Rivers interact with the former
96 draining into Lake Albert and the latter into the Victoria Nile (Onyutha et al. 2021) (Fig. 1).



97

98 **Fig. 1** Map the Lake Edward system and the neighbouring Lakes Albert and Victoria. Sample locations,
99 in which *Enteromius* was caught, of the four HIPE expeditions are given

100

101 *Enteromius* is the second most species-rich genus from the Lake Edward system, after *Haplochromis*
102 (Decru et al., 2019). Six species of *Enteromius* were historically reported from the Lake Edward system.
103 After morphological examination by Decru et al. (2019) of all the historical samples from the Lake
104 Edward system including the Congolese side, of which no genetic material is available, *E. neumayeri*
105 (Fischer, 1884) was removed from the species list. Two of the five remaining species belong to the
106 'soft-rayed barbs': *E. alberti* (Poll, 1939) and *E.* cf. *mimus* (Boulenger, 1912) (Maetens et al., 2020),
107 whereas *E. apleurogramma* (Boulenger, 1911), *E. kerstenii* (Peters, 1868) and *E. pellegrini* (Poll, 1939)
108 belong to the 'sawfin barbs'. The latter three species can be identified, respectively, by the
109 characteristic red fins and incomplete lateral line, a red dot on the operculum, and a black blotchy
110 pattern on the upper part of the flanks (Fig. 2). Decru et al. (2022) sequenced the COI gene and
111 identified seven lineages to occur within these three 'sawfin barbs' from the Lake Edward system: two
112 in both *E. apleurogramma* and *E. kerstenii*, and three in *E. pellegrini*. Genetic divergence was larger
113 than 2% in the COI gene, which is a general used consensus to delineate species (e.g. Hebert et al.,

2003; Ward et al., 2009; Pereira et al., 2013) and is mentioned in the structure of the FISH-BOL project (Ward, 2012). Therefore, these results suggest that each of the three species represent species complexes with a set of several species with extremely high similarities in morphology. The taxonomy of *E. pellegrini* in the system is problematic in view of its morphological similarities with *E. neumayeri* (Decru et al., 2019). The former species was described from Lake Kivu (DR Congo) and the latter from an affluent of the Ewaso Ng'iro River within the Lake Natron system (Kenya). In addition, one of the synonyms of *E. neumayeri*, *E. portali* (Boulenger, 1906), was described from the Lake Edward system.

Fig. 2 Overview of the reported 'safwin barbs' from the Lake Edward system: **a)** *E. apleurogramma* (HP3124), **b)** *E. kerstenii* (HP3123) and **c)** *E. pellegrini* (HP3127). All specimens were caught in the Mubuku river, Fort Portal-Mpondwe road in the Lake Edward system on 30 January 2018

In this study, we explored the diversity and evolutionary history of the 'sawfin barbs' from the Lake Edward system based on a combination of genetic (mitochondrial and nuclear markers) and morphometric approaches. In addition to Decru et al. (2022), where only the mitochondrial COI gene was used, we also sequenced, for a selection of specimens, the mitochondrial Cyt b gene and two nuclear genes: RAG1 and GH intron 2 following Schmidt et al. (2017). Furthermore, we performed morphometric analyses and further analysed the phylogenetic and morphometric relationships within

the three complexes including representatives from the adjacent systems of Lakes Albert and Victoria. We also investigated the temporal scale of the speciation events involved using a molecular clock approach. Finally, we aimed to place our results in a wider taxonomic, phylogenetic and geographic framework by including relevant data from GenBank of species of *Enteromius* from other river systems of the East Coast Province and the Nile Basin and from the neighbouring basin of the Congo.

2. Material and Methods

2.1. Taxon sampling

Four expeditions to the Lake Edward system took place at the Ugandan side of the system during wet (October 2016, March 2017, March 2019) and dry seasons (January 2018). It was not possible to sample the Congolese side of the system because of safety issues. Specimens of *Enteromius* were mainly caught in rivers using gill nets with different mesh sizes, a scoop net, and a backpack electrofisher. All specimens were euthanised with an overdose of clove oil. Some specimens were selected for molecular analyses and thus a tissue sample (fin clip) was taken and stored in 100% ethanol. All specimens were then fixed in formalin, rinsed with water, transferred to 70% denaturated ethanol and deposited in the collection of the Royal Museum for Central Africa, Tervuren. Some additional specimens were collected in the nearby regions of the systems of Lakes Albert and Victoria (Fig. 1). Information on the sample locations can be found in Table S1.

2.2. DNA extraction, PCR amplification and sequencing

In total, 275 DNA extractions of specimens of *Enteromius* with a serrated dorsal spine (i.e. 'sawfin barbs') from the Lake Edward system and neighbouring regions were used, of which 232 were already analysed in the study of Decru et al. (2022). We included 43 newly extracted samples from the Lake Edward system and neighbouring regions in the systems of Lakes Albert and Victoria using the standard protocol for human and animal tissue and cultured cells for the Nucleospin® Tissue Kit (Macherey-Nagel, Düren, Germany) (Table S1). The COI gene was isolated and amplified using the M13-tailed primer cocktail of Ivanova et al. (2007) and the protocol described by Decru et al. (2016). The amplification of the Cyt b gene was done with the primers L14724 (5'- GACTTGAAAAACCACCGTTG-3') and H15915 (5'-CTCCGATCTCCGGATTACAAGAC-3') (Xiao et al., 2001), with the following temperature profile: 94°C for 60 s (initial denaturation); 35 cycles of 94°C for 30 s, 54°C for 30 s, 72°C for 90 s; 72°C for 10 min (final extension). For the amplification of the nuclear marker RAG1, primers 2533F (5'-CTGAGCTGCAGTCAGTACCATAAGATGT-3') and 4090R (5'-CTGAGTCCTTGTGAGCTTCCATRAAYTT-3')

were used, following the protocol from López et al. (2004). GH intron 2 was amplified using primers GH23F (5'-TGTCGGTGGTGCTGGTCAGT-3') and GH148R (5'-TCCTTTCCGGTGGGTGCCTCA-3') from Mayden et al. (2009) and Schmidt et al. (2017), with the following PCR temperature profile: 94°C for 60 s (initial denaturation); 35 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 40 s; 72°C for 5 min (final extension). The amplified products were then visualised on a 1.2% agarose gel and purified using ExoSAP (Fermentas). Sequencing was executed bidirectionally by the company Macrogen.

2.3. Sequence alignment and phylogenetic analyses

Sequences were verified by eye in CodonCode Aligner v5.0.2 (CodonCode Corporation). Sequences were then aligned in MEGA v7.0.26 (Kumar et al., 2016) using MUSCLE (Edgar, 2004) and visually checked for amino acid substitutions and the presence of stop codons. The indels of GH intron 2 were coded in FastGap v1.2 (Borchsenius, 2009) by using the simple coding method (Simmons & Ochoterena, 2000). The gap characteristics were added to the end of the GH intron 2 alignment in PAUP* 4.0a (168) (Swofford, 2003). The length of the sequences was between 582 bp and 651 bp for COI, between 1099 bp and 1115 bp for Cytb, between 1448 bp and 1450 bp for RAG1 and between 216 bp and 240 bp for GH intron 2.

First, a maximum likelihood (ML) tree was reconstructed in IQ-TREE (Minh et al., 2020) for 243 COI sequences of *Enteromius* from the intensively sampled and well-defined Lake Edward system (Table S1). Sequences of two Asian small barbs, *Pethia ticto* (Hamilton, 1822) and *Hampala macrolepidota* Kuhl & Van Hasselt, 1823, were used as outgroups, following Ren & Mayden (2016) and Schmidt et al. (2017) (Table S1). DNA and codon models were tested separately using ModelFinder (Kalyaanamoorthy et al., 2017) and the best model was selected according to the lowest BIC value. The ML tree was constructed with the codon model MG+F3X4+G4 and 1000 ultrafast bootstrap replicates (UFBoot) (Hoang et al., 2018). The tree was rescaled with a factor 1/3 in R v4.3.1 (R Core Team, 2020) using R Studio v1.1.453 (RStudio Team, 2016) with package 'ape' v5.8 (Paradis & Schliep, 2019), because branch lengths under codon models represent the number of nucleotide substitutions per codon site, which is three times longer than under DNA models. The resulting lineages were named after their geographical distribution. The epithet "sp." was used, following Schmidt et al. (2017; 2019), to highlight likely new species with most similarities to one of the nominal species. A K2P (Kimura 2-P) distance matrix for the lineages was calculated in MEGA 7.0.26 and the presence of a barcoding gap was examined.

In a second step, A TCS haplotype network (Clement et al., 2000) was made for each species complex in PopART v1.7 (Leigh & Bryant, 2015) based on the COI dataset from Lake Edward, including 32

additional sequences of *Enteromius* that were *ad hoc* collected in the associated systems of Lakes Albert and Victoria (Table S1). An additional K2P distance matrix was calculated in MEGA 7.0.26 with the inclusion of these extra samples. Thirty samples of *Enteromius* from the Lake Edward, Albert and Victoria systems were selected , representing all lineages, to create a ML tree in IQ-TREE using a partition model (Chernomor et al., 2016) including the mitochondrial markers COI and Cyt b and the nuclear markers RAG1 and GH intron 2. The following models were selected, using ModelFinder: the codon models MG+F3X4+G4, MG+F3X4+R3, SCHN05+FU+G4 and the DNA model HKY+F for COI, Cyt b, RAG1 and GH intron 2, respectively. *Pethia ticto* and *H. macrolepidota* were included as outgroups and 1000 ultrafast bootstrap replicates were used.

Finally, a phylogenetic study, using a wide geographic context, was performed including sequences from GenBank for COI and Cytb from neighbouring basins of the Nile and the Congo (Table S1). Identifications are those as on GenBank. For the COI gene, we additionally included seven sequences of *E. pellegrini* and four of *E. radiatus* (Peters, 1853) from Lake Kivu and the Lowa Basin (Kisekelwa et al., 2022) and one of *E. neumayeri* from the Southern Ewaso Ng'iro, Kenya (Schmidt et al., unpublished material), which were not deposited on GenBank (Table S1). For each marker, a ML tree was made in IQ-TREE using ModelFinder and 1000 ultrafast bootstrap replicates. The codon models MG+F3X4+G4 and MG+F3X4+R3 were selected for COI and Cyt b, respectively, using ModelFinder. The resulting tree was rescaled with factor 1/3. Sequences of *P. ticto* and *H. macrolepidota* were used as outgroups (Table S1). All ML trees were visualized using FigTree v1.1.4 (Rambout, 2018).

2.4. Molecular clock

A molecular clock analysis was conducted using BEAST v1.8.4 (Drummond et al., 2012). One sequence per lineage was selected for COI, Cyt b and RAG1. The gene GH intron 2 was not included, because of the rather short length of the intron. *Puntigrus tetrazona* (Bleeker, 1855) was selected as an outgroup (Table S1). The alignments of COI, Cyt b and RAG1 were analysed as different partitions with different substitution models. The models for each alignment were individually tested in JModelTest2 (Darriba et al., 2012) according to the lowest BIC value. The substitution model for the two mitochondrial markers was HKY+Γ and K80+ Γ for the nuclear marker. A birth-death model and an uncorrelated lognormal relaxed clock (UCLN) were used. To time-calibrate our phylogenetic tree, we used the divergence dates between *P. tetrazona* and the small African diploid barbs: 30.1 Mya (Million years ago) as found in Ren & Mayden (2016) with a standard deviation of 2.1 and a normal prior distribution. This calibration is in accordance with the current understanding of the biogeographical scenario regarding the origin of *Enteromius*, where the ancestral populations of small barbs dispersed from the

Oriental into the Afrotropical region between 30.5 and 23 million years ago (Lavoué et al., 2020). Regarding the Markov chain Monte Carlo (MCMC) parameters, a chain length of 10 million states, sampling every 1000 iterations and thus resulting in a total of 10000 sampled trees was used. The mixing and convergence of the MCMC chains were visually checked using Tracer v1.7.1. (Rambaut et al., 2018) by ensuring that all statistics have effective sampling size (ESS) above 200. We finally constructed a maximum clade credibility (MCC) tree using TreeAnnotator v1.8.4 (Drummond & Rambaut, 2007) with a 25% burn-in. MCC trees were visualized using FigTree v1.1.4.

2.5. Morphometric analyses

For each of the genetic lineages identified, a selection of specimens from the Lake Edward system and neighbouring regions was examined morphologically (Table S1). In total, 69 specimens were examined: 15 of the *E. apleurogramma* complex, 22 of the *E. kerstenii* complex and 32 of the *E. pellegrini* complex (Table S1). The four type specimens of *E. portali* (Boulenger, 1906) were also included, because this junior synonym of *E. neumayeri* was described from the Lake Edward system (Table S1). For each specimen, 24 measurements were taken with dial callipers, and 18 meristics were counted based on Bamba et al. (2011) and modified by Maetens et al. (2020). Measurements and meristics were analysed separately with principal component analyses (PCA). Measurements were log-transformed and their covariance matrix analysed. The first axis of this analysis is interpreted as a proxy for size (Bookstein et al., 1985). For the PCA on meristics, the correlation matrix of the raw data was used. The data of the barbels were excluded from the analyses, because they are fragile and break off easily (Maetens et al., 2020). Invariable meristics were excluded from the PCA. The analyses were done in R v4.3.1 using R Studio v1.1.453 and the packages 'vegan' v2.5-6 (Oksanen et al., 2019) and 'factoextra' v1.0.7 (Kassambara & Mundt, 2020).

3. Results

3.1. Three species complexes within the Lake Edward system

We examined the specimens from the Lake Edward system by assessing their COI sequences and morphology. The ML tree of COI (Fig. 3a) showed three species complexes with a total of seven lineages with genetic divergences larger than 2% (Table 1). Two lineages could be assigned to the species complex of *E. apleurogramma*: *E.* sp. 'apleurogramma highland' and *E.* sp. 'apleurogramma lowland', another two to the species complex of *E. kerstenii*: *E.* sp. 'kerstenii highland' and *E.* sp. 'kerstenii lowland', and the remaining three belonged to the species complex of *E. pellegrini*: *E.* sp. 'pellegrini

highland', *E.* sp. 'pellegrini lowland' and *E.* sp. 'pellegrini Ishasha upstream'. Highland specimens were generally caught at higher altitudes in the eastern part of the system except for five specimens of *E.* sp. 'kerstenii highland' and three specimens of *E.* sp. 'pellegrini highland', which were caught at lower altitudes in the centre of the system close to the lakes (Fig. 4b, c). Specimens of *E.* sp. 'pellegrini Ishasha upstream' were caught in the upper reaches of the Ishasha River, in the extreme southern part of the Edward system, close to the hydrological boundary with the Lake Victoria system (Fig. 4c). One aberrant specimen (HP2674) is nested within *E.* sp. 'pellegrini highland', although with a genetic distance of 1.97% compared to the other specimens within this lineage (Fig. 3a). This specimen was caught close to Lake Edward, in sympatry with specimens of *E.* sp. 'pellegrini lowland' (Fig. 4c). A barcoding gap was found between the intra- and interspecific distances when excluding the aberrant specimen HP2674 (Fig. S1a, Table 1). With the inclusion of specimen HP2674, no barcoding gap could be observed (Fig. S1b, Table 1).

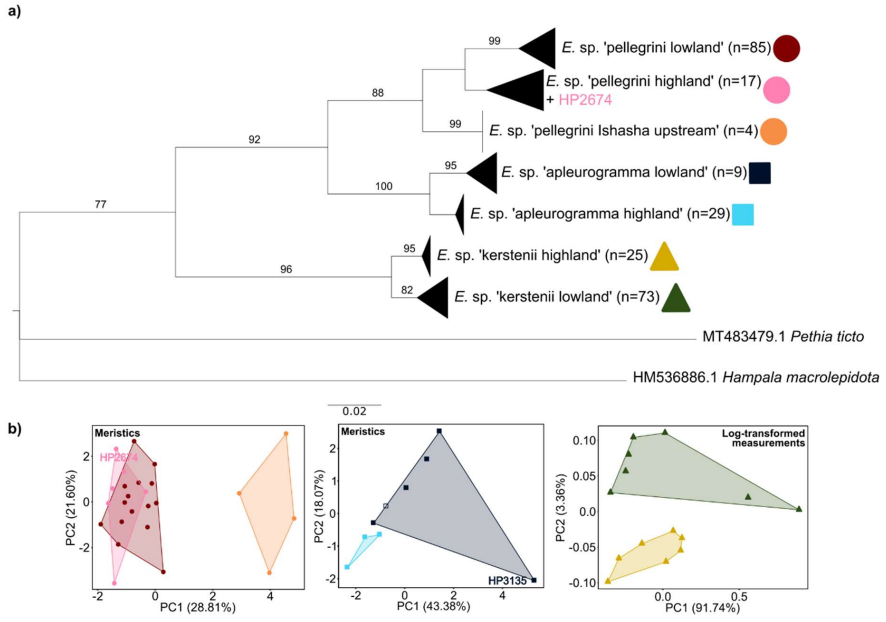


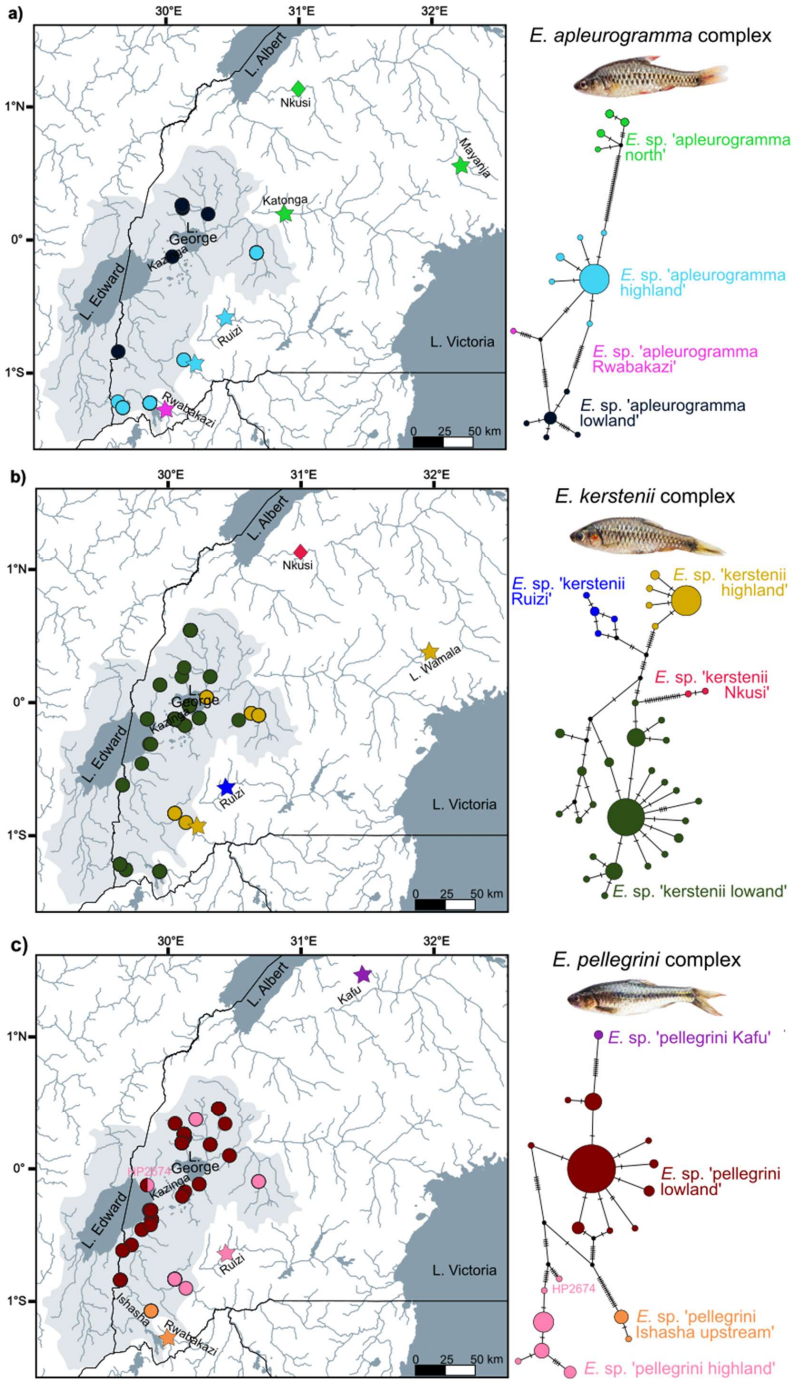
Fig. 3 a) ML tree of 243 COI sequences retrieved of the 'sawfin barbs' from the Lake Edward system with 1000 UFBoot replicates. Bootstrap values (>75%) are given above the branches. **b)** PCAs on log-transformed measurements and meristics of the 'sawfin barbs' from the Lake Edward system, filled symbols represent one specimen, open symbols represent two specimens. Lineages with less than 2% genetic divergence were collapsed. See text for special position of HP2674

Table 1 Overview of the minimum and maximum within group and between group distances (in %), using the K2P distance model, of the seven lineages of the sawfin barbs from the Lake Edward system. The dotted lines separate the three species complexes: *E. apleurogramma* complex (**1.** *E.* sp. 'apleurogramma highland', **2.** *E.* sp. 'apleurogramma lowland'), *E. kerstenii* complex (**3.** *E.* sp. 'kerstenii highland', **4.** *E.* sp. 'kerstenii lowland') and *E. pellegrini* complex (**5.** *E.* sp. 'pellegrini highland', **6.** *E*. sp. 'pellegrini lowland', **7.** *E*. sp. 'pellegrini Ishasha upstream'). Genetic distances with the exclusion of the aberrant specimen HP2674 (group 5) are highlighted

	Within group	Between group					
		1	2	3	4	5	6
1	<0.01-0.31						
2	<0.01-1.10	2.19-3.18					
3	<0.01-0.31	12.40-13.84	11.64-13.00				
4	<0.01-1.72	12.02-14.06	11.26-13.55	2.19-3.33			
5	<0.01-2.36	7.89-9.97	8.05-9.68	12.09-13.84	11.65-14.44		
	<0.01-1.09	7.89-9.18	8.05-9.68	12.09-12.48	11.65-13.26		
6	<0.01-1.61	7.88-9.55	8.77-10.39	12.17-14.14	12.17-14.36	2.90-4.88	
						2.90-4.82	
7	<0.01	8.25-8.98	8.05-8.41	13.22-13.62	12.64-14.21	4.98-5.50	4.32-5.83
						4.99-5.50	

Within each species complex, specimens from the various lineages could be distinguished based on morphometric analyses either with meristics or log-transformed measurements, except within the *E. pellegrini* complex (Fig. 3b). Based on a PCA on meristics, we found that specimens of *E.* sp. 'pellegrini Ishasha upstream' were completely separated from the highland and lowland specimens on PC1 with the number of scales between the lateral line and the base of the pelvic fin, the number of pre-dorsal scales and the number of scales between the base of the dorsal fin and the lateral line as the most important loadings (Fig. 3b, Table S2), while the morphospace of the specimens of the highland and lowland populations almost completely overlapped. An additional PCA including only meristics of these highland and lowland specimens did not result in a better discrimination (not illustrated). Based on a PCA on log-transformed measurements, the three lineages largely overlapped (not illustrated). The lineages within the *E. apleurogramma* complex could be distinguished based on a PCA on meristics (Fig. 3b, Table S2) with the number of scales between the lateral line and the base of the pelvic fin, the number of branched pectoral-fin rays and the number of gill rakers on the upper part of the first gill arch, and the number of lateral line scales (with and without caudal fin scales) as most important loadings on PC1. The number of branched pelvic-fin rays and the total number of gill rakers on the first gill arch were the most important loadings on PC2. One outlier within *E.* sp. 'apleurogramma lowland' (HP3135) differed from the other specimens within this lineage by a different number of gill rakers and branched pectoral fin rays. No clear distinction between lineages was found based on log-transformed measurements (not illustrated). The two lineages of the *E. kerstenii* complex slightly overlapped in a

PCA on meristics (not illustrated). The two lineages could be distinguished on PC2 of a PCA on log-transformed measurements (Fig. 3b, Table S2) with the interorbital width, the pectoral fin length and the body depth as the most important loadings on PC2.


a)
30°E
31°E
32°E
1°N
0°
1°S
L. Albert
Nkusi
Mayanja
Katonga
L. George
Kazinga
L. Edward
Ruizi
L. Victoria
Rwabakazi
0 25 50 km
E. apleurogramma complex
E. sp. 'apleurogramma north'
E. sp. 'apleurogramma highland'
E. sp. 'apleurogramma Rwabakazi'
E. sp. 'apleurogramma lowland'
b)
30°E
31°E
32°E
1°N
0°
1°S
L. Albert
Nkusi
L. Wamala
L. George
Kazinga
L. Edward
Ruizi
L. Victoria
0 25 50 km
E. kerstenii complex
E. sp. 'kerstenii Ruizi'
E. sp. 'kerstenii highland'
E. sp. 'kerstenii Nkusi'
E. sp. 'kerstenii lowand'
c)
30°E
31°E
32°E
1°N
0°
1°S
L. Albert
Kafu
L. George
HP2674
Kazinga
L. Edward
Ruizi
Ishasha
Rwabakazi
L. Victoria
0 25 50 km
E. pellegrini complex
E. sp. 'pellegrini Kafu'
E. sp. 'pellegrini lowland'
HP2674
E. sp. 'pellegrini Ishasha upstream'
E. sp. 'pellegrini highland'

313

Fig. 4 TCS haplotype network of the COI sequences and geographical distribution per species complex: **a)** *E. apleurogramma*, **b)** *E. kerstenii* and **c)** *E. pellegrini*. The Lake Edward system is highlighted in grey. Specimens from the drainages of Lakes Albert and Victoria are indicated with a diamond and a star respectively

3.2. Lineages of the three complexes in the adjacent systems of Lakes Albert and Victoria

We expanded the genetic and morphometric analyses of the 'sawfin barbs' from the Lake Edward system by including a second mitochondrial marker and two nuclear markers and by examining additional specimens from Lakes Albert and Victoria. For each of the species complexes, we found some of the specimens from the Lake Victoria system clustering with the lineages from higher regions from the Lake Edward system, based on the TCS haplotype network of COI sequences (*E.* sp. 'apleurogramma highland', *E.* sp. 'kerstenii highland', *E.* sp. 'pellegrini highland' and *E.* sp. 'pellegrini Ishasha upstream'; Fig. 4). Other specimens from the systems of Lakes Victoria and Albert formed distinct lineages within the species complexes: *E.* sp. 'apleurogramma North', *E.* sp. 'apleurogramma Rwabakazi', *E.* sp. 'kerstenii Nkusi', *E.* sp. 'kerstenii Ruizi' and *E.* sp. 'pellegrini Kafu' (Fig. 4). The minimum and maximum within-group and between-group genetic distances are given in Table 2. The existence of the 12 distinct lineages was confirmed in the concatenated tree with mitochondrial and nuclear markers (Fig. 5).

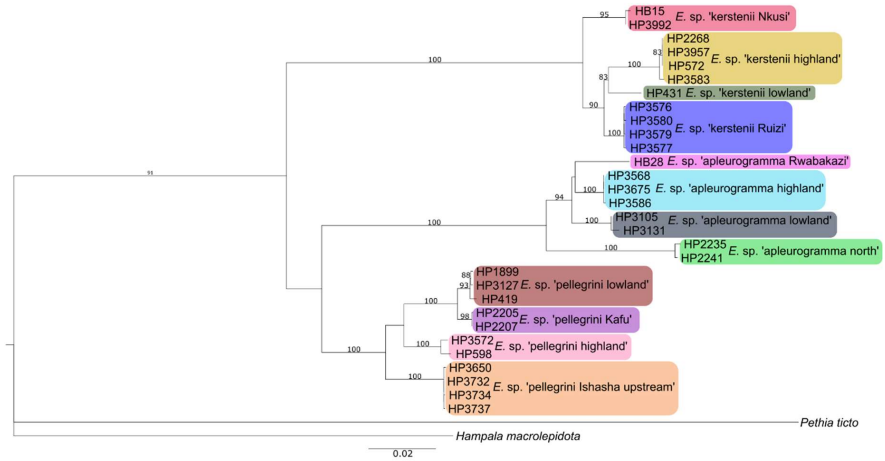


Fig. 5 A ML tree of the concatenated mitochondrial genes COI and Cyt b and the nuclear genes RAG1 and GH intron 2 of the specimens of *Enteromius* with a serrated dorsal spine from the Lake Edward system and the associated regions of Lakes Albert and Victoria. Bootstrap values (>75%) are given above the branches

Table 2 Overview of the within group and between group distances (in %), using the K2P distance model, of the twelve lineages from the Lake Edward system, including specimens from Lakes Albert and Victoria. The dotted lines separate the three species complexes: *E. apleurogamma* complex (**1.** *E.* sp. 'apleurogramma highland', **2.** *E.* sp. 'apleurogramma lowland', **3.** *E.* sp. 'apleurogramma Rwabakazi', **4.** *E.* sp. 'apleurogramma north'), *E. kerstenii* complex (**5.** *E.* sp. 'kerstenii highland', **6.** *E.* sp. 'kerstenii lowland', **7.** *E.* sp. 'kerstenii Nkusi', **8.** *E.* sp. 'kerstenii Ruizi'), *E. pellegrini* complex (**9.** *E.* sp. 'pellegrini highland', **10.** *E.* sp. 'pellegrini lowland', **11.** *E.* sp. 'pellegrini Ishasha upstream', **12.** *E.* sp. 'pellegrini Kafu'). Genetic distance with the exclusion of the aberrant specimen HP2674 (group 9) are highlighted

	Within group	Between group										
		1	2	3	4	5	6	7	8	9	10	11
1	<0.01-0.31											
2	<0.01-1.10	2.19-3.18										
3	/	1.56-1.72	2.51-3.53									
4	<0.01-0.93	6.16-7.41	6.48-7.81	7.01-7.72								
5	<0.01-0.31	12.40-13.84	11.64-13.00	11.81-12.19	15.08-15.89							
6	<0.01-1.72	12.02-14.06	11.26-13.55	11.06-12.12	14.48-15.28	2.19-3.34						
7	0.15	12.03-13.21	11.09-12.42	11.45-11.64	14.90-15.51	3.33-3.82	2.84-4.16					
8	<0.01-0.31	12.59-13.82	11.83-13.38	12.00-12.37	14.28-15.26	1.87-2.51	1.56-2.68	3.17-3.65				
9	<0.01-2.36	7.89-9.97	8.05-9.68	8.78-9.86	9.31-10.59	12.09-13.84	11.65-14.44	11.92-14.05	12.29-14.21			
	<0.01-1.09	7.89-9.18	8.05-9.68	8.78-9.33	9.31-10.42	12.09-12.48	11.65-13.26	11.92-13.08	12.29-12.85			
10	<0.01-1.61	7.88-9.55	8.77-10.39	8.58-10.03	8.76-10.78	12.17-14.14	12.17-14.36	12.40-14.16	13.23-14.80	2.90-4.82		
										2.91-4.82		
11	<0.01-0.15	8.25-9.16	8.05-8.58	8.78-8.95	8.59-9.12	13.22-13.80	12.64-14.39	13.64-14.01	13.62-14.00	4.98-5.66	4.32-5.99	
										4.98-5.66		
12	0.15	9.69-10.78	9.48-10.02	10.23-10.41	7.54-10.58	12.66-13.22	12.83-14.39	14.05-14.44	13.84-14.39	4.48-5.50	1.40-2.60	5.32-5.65
										4.83-5.50		

We found, for the *E. apleurogramma* complex, that *E.* sp. 'apleurogramma north' was separated from the other lineages on PC1 of the PCA on meristics (Fig. 6a), with the number of lateral line scales and pre-dorsal scales as most important loadings (Table S3). However, these results are based on the morphological examination of only one specimen of *E.* sp. 'apleurogramma north' because of the poor condition of the other specimens of this lineage. *Enteromius* sp. 'apleurogramma lowland' overlapped slightly with *E.* sp. 'apleurogramma highland'. This overlap was caused by the additional specimens of *E.* sp. 'apleurogramma highland' from the system of Lake Victoria that were included in this analysis. The one specimen of *E.* sp. 'apleurogramma Rwabakazi' was situated within the polygon of *E.* sp. 'apleurogramma highland'. A similar result was obtained when *E.* sp. 'apleurogramma north' was

359 excluded from the analysis (not illustrated). No separation was found based on a PCA on the log-
360 transformed measurements (not illustrated). Based on meristics, the different lineages within the *E.*
361 *kerstenii* complex could not be separated (not illustrated). However, all lineages could be distinguished
362 from each other on a combination of the second and third axis of the PCA on the log-transformed
363 measurements (Fig. 6b), except for one overlapping specimen of *E.* sp. 'kerstenii highland' with *E.* sp.
364 'kerstenii lowland'. For PC2, the most important loadings were body depth and anal-fin length. For
365 PC3, these were the anal-fin-base length and the post-dorsal-fin distance (Table S3). In a PCA on the
366 log-transformed measurements of all specimens within the *E. pellegrini* complex (Fig. 6c), specimens
367 of *E.* sp. 'pellegrini Ishasha upstream' were separated from all other lineages PC1 with the number of
368 scales between the dorsal fin and the lateral line and the number of pre-dorsal scales as most
369 important loadings (Table S3). *Enteromius* sp. 'pellegrini highland', *E.* sp. 'pellegrini lowland' and *E.* sp.
370 'pellegrini Kafu' were largely overlapping. Based on a PCA of the log-transformed measurements, no
371 separation was found (not illustrated). We found the type specimens of *E. portali* to be separated from
372 all the lineages within the *E. pellegrini* complex on PC2 of a log-transformed PCA with the dorsal-fin
373 length and body depth as main loadings (Fig. 6d, Table S3). In summary, all lineages can be separated
374 morphologically except for *E.* sp. 'apleurogramma highland' and *E.* sp. 'apleurogramma Rwabakazi'; *E.*
375 sp. 'apleurogramma highland' and *E.* sp. 'apleurogramma lowland'; and *E.* sp. 'pellegrini highland', *E.*
376 sp. 'pellegrini lowland' and *E.* sp. 'pellegrini Kafu'.

377

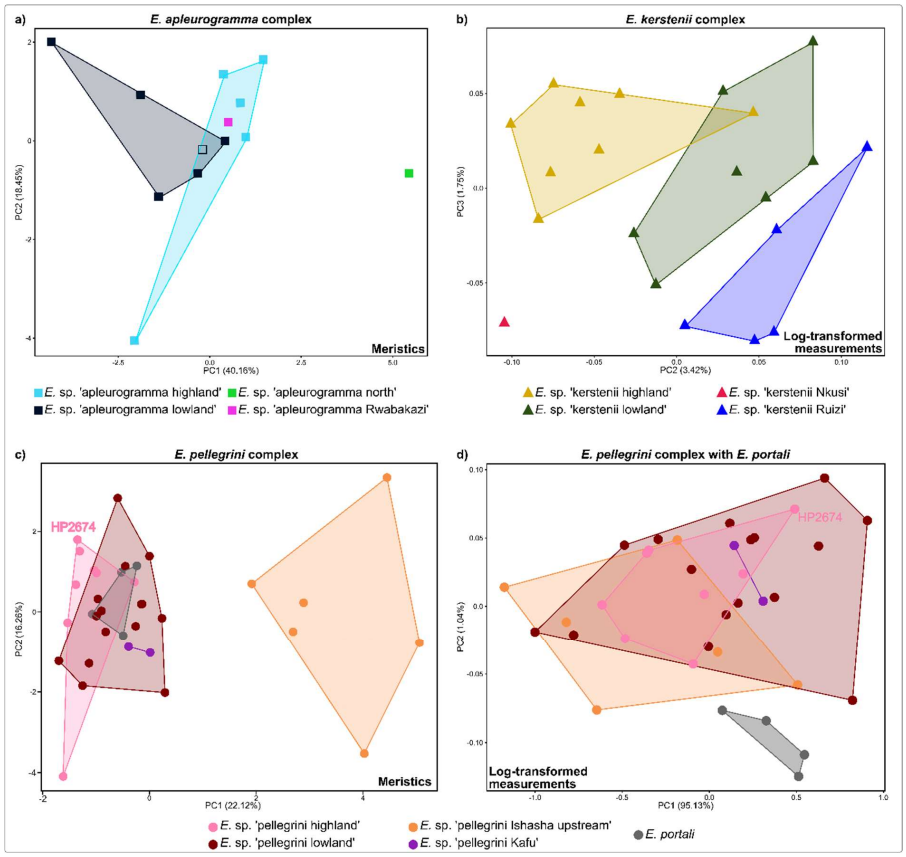


Fig. 6 Scatterplots of PCAs on specimens from the systems of Lakes Edward, Albert and Victoria with a**)** PC2 against PC1 of the PCA on 12 meristics of 14 specimens of the *E. apleurogramma* species complex, **b)** PC3 against PC2 of the PCA on 24 log-transformed measurements of 21 specimens of the *E. kerstenii* species complex, **c)** PC2 against PC1 of the PCA on 14 meristics of 31 specimens of the *E. pellegrini* species complex, and **d)** PC2 against PC1 of the PCA on 24 log-transformed measurements of 32 specimens of the *E. pellegrini* species complex and the four syntypes of *E. portali*. Filled symbols represent one specimen, open symbols represent multiple specimens

3.3. Ancient lineages of 'sawfin barbs' from the Lake Edward system and adjacent systems

We performed a molecular clock on a selection of specimens to get an estimated time frame of the origin of the lineages of 'sawfin barbs' from the Lake Edward system (Fig. 7). The analysis conducted in BEAST revealed that the split between the different complexes happened during the Miocene (Fig. 7, Table 3). The time to the most recent common ancestor (TMRCA) for the complexes of *E. apleurogramma* and *E. pellegrini* were older (Pliocene) than the TMRCA for the complex of *E. kerstenii*,

which is around the Pliocene-Pleistocene. Most of the divergence within the species complexes occurred during the Pliocene and the Pleistocene (Fig. 7).

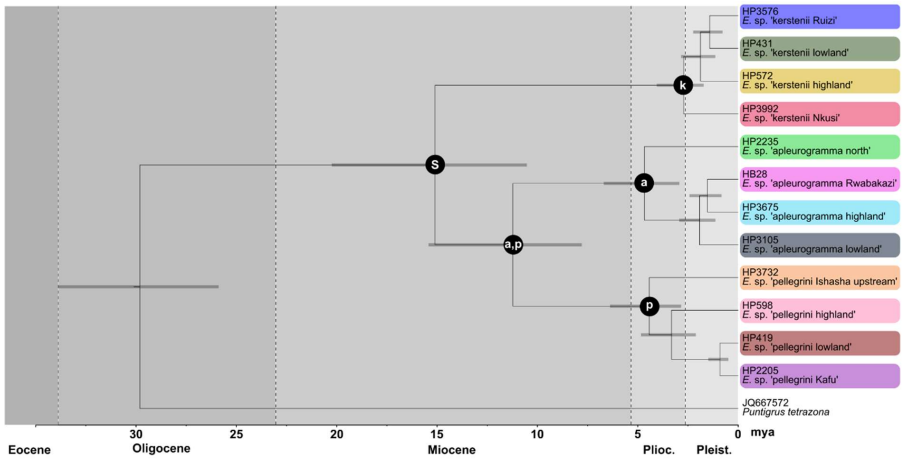


Fig. 7 MCC tree of the different markers (COI, Cyt b, and RAG1) with a, k and p representing the most recent common ancestor for the *E. apleurogramma* complex, *E. kerstenii* complex and *E. pellegrini* complex respectively. The most recent common ancestor for *E. apleurogramma* – *E. pellegrini* and *E. kerstenii* – *E. pellegrini* are indicated with a-k and a-p respectively. The most recent common ancestor for the 'sawfin barbs' from the region examined is indicated with S

Table 3 Overview of the median times to the most recent common ancestor (Mya) and the 95% highest posterior density (HPD) boundaries.

	Median	Lower (95% HPD)	Upper (95% HPD)
'Sawfin barbs' (S)	15.10	10.53	20.25
***E. apleurogramma* – *E. pellegrini* (a, p)**	11.22	7.79	15.43
***E. apleurogramma* complex (a)**	4.67	2.93	6.70
***E. kerstenii* complex (k)**	2.72	1.72	4.06
***E. pellegrini* complex (p)**	4.43	2.84	6.38

3.4. An overarching phylogenetic and geographical framework

Within the larger phylogenetic framework that included sequences of 'sawfin barbs' from other regions of the East Coast, the Nile Basin and from the neighbouring drainage basins of the Congo, we found that within each complex, additional lineages and sometimes specimens with a doubtful GenBank identification were found (Fig. 8, S2, S3, Table S4). Based on COI, the *E. kerstenii* complex is

situated within a clade of mainly *E. miolepis*-like lineages from the Congo Basin. The specimens identified as *E. kerstenii* from the coastal basins of Tanzania (close to type locality of *E. kerstenii*) form a monophyletic group with the *E. kerstenii* complex from the Lake Edward system. The complexes of *E. apleurogramma* and *E. pellegrini* form sister clades. Specimens of *E.* sp. 'pellegrini lowland' and *E.* sp. 'pellegrini Kafu' clustered together with specimens identified as *E. neumayeri* from the southern Ewaso Ng'iro River in Kenya (type locality of *E. neumayeri*). Specimens of the complex of *E. pellegrini* form a monophyletic group with specimens identified as *E. pellegrini*, *E.* cf. *pellegrini* and *E. trinotatus* from the Congo Basin (Epulu and Ituri). For Cyt b, similar patterns were observed (Fig. S3, Tables S1, S4). Additional specimens identified as *E. apleurogramma* from Lakes Victoria, Kanyaboli, the Uriri dam and the Athi River form a monophyletic clade with the specimens of *E.* sp. 'apleurogramma north'. Most of the specimens identified on GenBank as *E. kerstenii* form a monophyletic group with the *E. kerstenii* complex of the Lake Edward system and neighbouring regions. However, several specimens identified as *E. kerstenii* are situated within the clades of *E. pellegrini*, *E. apleurogramma* and *E. nyanzae* (see also Schmidt et al., 2017).

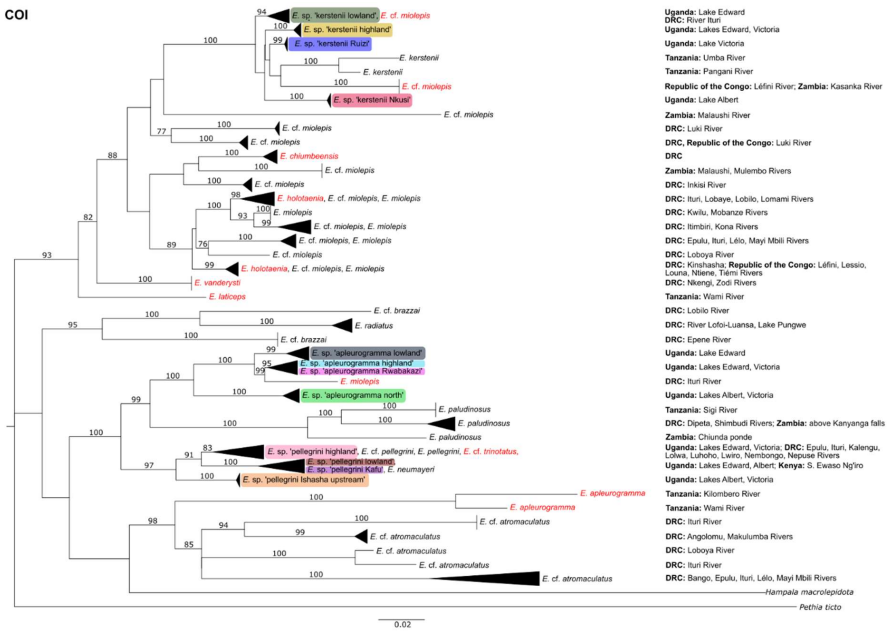


Fig. 8 ML tree of the mitochondrial COI gene of the 'sawfin barbs' from the Lake Edward system and the neighbouring basins of the Nile, Congo and other regions in the East Coast ichthyofaunal provinces. Bootstrap values (>75%) are given under the branches. Specimens from our study are highlighted,

specimens with a doubtful GenBank identification are in red. Lineages with less than 2% genetic divergence were collapsed

4. Discussion

4.1. Species diversity of the 'sawfin barbs' from the Lake Edward system and neighbouring regions

The combination of genetic divergences and differences in morphology (multivariate analyses) can be a strong method to delineate species in *Enteromius* (e.g. Van Ginneken et al., 2017; Schmidt et al., 2018; Englmaier et al., 2020; Kambikambi et al., 2021). Using the criteria of Seehausen et al. (1998) that species can be delineated when they are supported by two characters that are presumed to be genetically independent, all lineages from the Lake Edward system, excluding the specimens from Lakes Albert and Victoria, in the complexes of *E. kerstenii* and *E. apleurogramma* can be considered distinct biological species. Indeed, all of them can be separated with morphological and genetic data with the following exception. Within the *E. pellegrini* complex, the lineages of *E*. sp. 'pellegrini highland' and *E.* sp. 'pellegrini lowland' could not be distinguished using our morphological characters, although they were genetically distinct. The same held for some of the lineages occurring outside of the Lake Edward system. *Enteromius* sp. 'apleurogramma highland' and *E.* sp. 'apleurogramma Rwabakazi' had a genetic distance between 1.59% and 1.72% (COI) and could not be separated morphologically. Within the *kerstenii* complex, specimens of *E.* sp. 'kerstenii lowland' and *E.* sp. 'kerstenii Ruizi' had a small genetic distance (1.56-2.68%, COI), but could be easily distinguished morphologically. Specimens of *E.* sp. 'pellegrini Kafu' overlapped morphologically with the specimens of *E.* sp. 'pellegrini lowland' with a genetic distance between 1.40% and 2.60% (COI). Nevertheless, for the discussion, we consider the twelve lineages of *Enteromius* from the Lake Edward system and the neighbouring regions as distinct biological species within three species complexes. The nuclear support is a strong evidence to consider the lineages as distinctive species, because of the lower evolutionary rate in nuclear DNA (Brown et al., 1979). We found one aberrant specimen (HP2674), which was genetically close to *E.* sp. 'pellegrini highland' based on COI and living in sympatry with *E.* sp. 'pellegrini lowland'. In a follow-up study where we are studying speciation patterns in the *E. pellegrini* complex via whole genome analyses, this specimen, based on its nuclear DNA, clustered together with specimens of *E.* sp. 'pellegrini lowland'. This may indicate a possible event of hybridisation. In addition, we examined the types of *E. portali*, a species described from the Lake Edward system, and morphologically very similar to *E. pellegrini* but currently a synonym of *E. neumayeri* (Decru et al., 2019). The morphological differences we found between these types and the species of the *E. pellegrini* complex suggest that the types of *E. portali* represent an additional valid species of *Enteromius* within the Lake Edward system. At present, no additional specimens, nor genetic information of this nominal species are available to verify this.

During recent expeditions, *Enteromius* specimens were caught near the type locality of *E. portali* (Fort Portal, Uganda) but none clustered with the types in the PCA on morphometrics. Unfortunately, no genetic material was available from the Congolese side of the Lake Edward system. This side of the lake is characterised by rocky shorelines and deep-water habitats. Decru et al. (2019) suggest a possible undetected diversity in these less common habitats. While specimens of *Enteromius* can be found along the lake's shoreline, they are mainly riverine. Therefore, the rocky and deep habitats, important for lacustrine species, may be less important for *Enteromius*. Nevertheless, the understudied rivers on the Congolese side of the system might reveal additional species within the species complexes already described from the Ugandan side.

Our results are reminiscent to the results from studies on *Enteromius* for the neighbouring Congo Basin, where four 'a priori' species turned out to represent 23 genetic lineages with a genetic divergence between 1.75% to almost 20% in the COI gene (Van Ginneken et al., 2017). Van Ginneken et al. (2017) also found small morphometric differences upon multivariate analyses for most of the lineages. We also included nuclear genes in the analyses, as did Schmidt et al. (2017, 2019), who found unrecognised diversity within *Enteromius* in Kenyan rivers and in West Africa. Unrecognised diversity was also found in South Africa (Chakona et al., 2015; Kambikambi et al., 2021). This shows that our results can likely be extrapolated to other regions in Africa, showing that species diversity *Enteromius* is highly underestimated and the genus may become the most species-rich fish genus in the world. The discrepancy between genetics and morphology is not unique in African freshwaters. A high genetic divergence and a near morphological stasis also occurs in other taxa. It was found for two populations of the monotypic butterfly fish *Pantodon buchholzi* Peters, 1876 (15.2% in mitogenome) (Lavoué et al., 2011). An opposite pattern is found in *Haplochromis*, the species of which are known to have low genetic divergences and large morphological differences (Verheyen et al., 2003; McGee, 2020). Outside of the African continent, a similar scenario of that of *Enteromius* can be found in the Asian small barbs, *Pethia* Pethiyagoda, Meegaskumbura & Maduwage, 2012. *Enteromius* and *Pethia* both belong to the tribe Smiliogastrini within Cyprinidae. Sudasinghe et al. (2021) found a genetic divergence (Cyt b) ranging between 0.8 and 8.4% between five species of *Pethia* of which only one species could be morphologically distinguished from the other four.

4.2. Geographical and possible evolutionary patterns

The Lake Edward system and the adjacent systems include marshy areas where the boundaries of the different systems are not well marked and where fish can possibly migrate between the different systems (Doornkamp & Temple, 1966; Greenwood, 1973; Decru et al., 2019). In each of the three

497 species complexes, we found specimens from the Lake Victoria system that seem to be conspecific to
498 the highland species from the Lake Edward system, which were caught in the eastern part of the
499 system close to the watershed with the Lake Victoria system. Within the Lake Edward system, each
500 species complex also includes a lowland species, predominantly caught at lower altitudes close to
501 Lakes Edward and George (Fig. 4). The geographical patterns of the lowland and highland species
502 suggest that the lowland species are endemic to the Lake Edward system with a narrow distribution
503 close to the lakes, while the broader geographical distribution of the highland species is probably the
504 result of a dispersal event between the Lake Edward system and the system of Lake Victoria. This
505 hypothesis is strengthened by the fact that many rivers in this part of Uganda have an east-west
506 division, with one part flowing to the east and the other to the west. These rivers are embedded in a
507 papyrus swamp where the direction of the flow cannot be determined (Beadle, 1974). The best-known
508 example hereof is the Katonga River (Fig. 1), which used to fully flow into Lake George via the Mpanga
509 River. Due to uplifting in this area during the mid-Pleistocene, the Katonga River now flows in two
510 directions from a swampy area (Beadle, 1974; Reardon & Chapman, 2009). Two additional such
511 swampy areas can be found along the eastern watershed of the Birira River, which flows into Lake
512 Edward via the Ntungwe River (Fig. 1). The first swampy region is situated at the headwaters of the
513 Ruizi River, which flows into a large swampy area that includes the Koki Lakes, which are drained by
514 the Kibali River towards Lake Victoria (Beadle, 1974; Odada et al., 2006; Decru et al., 2019). We found
515 specimens caught in the Ruizi River (Lake Victoria) which genetically clustered within the highland
516 specimens of *E. apleurogramma* and *E. pellegrini* (Fig. 4a, c). A second swampy region can be found at
517 the divide of the Birira River (Lake Edward system) and the Rufua-Kagera River (Fig. 1) (Lake Victoria)
518 (Beadle, 1974). Also for this swampy area we found specimens from the Lake Victoria system which
519 clustered genetically within the highland specimens of *E. apleurogramma* and *E. kerstenii* (Fig. 4a, b).
520 A possible fourth connection was found between the Ishasha River (Lake Edward system) and the
521 Rwabakazi River (Fig. 1) (Lake Victoria). Both rivers originate in adjacent wetland habitats (Kasangaki,
522 2006; Mbalassa et al., 2015; Mukasa-Tebandeke et al., 2020). For *E.* sp. 'pellegrini Ishasha' we found
523 four specimens caught in the Ishasha River (Lake Edward system) and two specimens in the Rwabakazi
524 River (Lake Victoria). However, within the *E. apleurogramma* complex, specimens caught in the Ishasha
525 River, i.e. *E.* sp. 'apleurogramma highland', differed genetically from those caught in the Rwabakazi
526 River, i.e. *E.* sp. 'apleurogramma Rwabakazi'. The presence of at least four swampy areas is testimony
527 of connections that make recent dispersal of species possible and those in the past between the two
528 systems in wetter periods. Between the systems of Lakes Albert and Victoria, swampy areas are also
529 present where possible migration of fishes could take place during wet periods. The best known
530 example hereof is the connection between the Nkusi River (Lake Albert) and the Kafu River (Lake
531 Victoria), which share a watershed where water flows in both directions (Fig. 1) (Beadle, 1974).The

phylogenetic patterns of *Enteromius* can be compared with the phylogeny of *Clarias gariepinus* (Buchell, 1822) presented by Van Steenberge et al. (2020). They traced the origin of *C. gariepinus* back to the East Coast province (4.8-1.65 Mya), which is similar to the interspecific divergences observed in *Enteromius* (4.67-2.72 Mya). Future research is needed to compare the distribution patterns of the different species of *Enteromius* with those of *C. gariepinus*. Consequently, *Enteromius* can become a test case for a general evolutionary scenario for African riverine fishes.

In contrast to the large body of work conducted on speciation and diversification in the species-rich East African cichlids (reviewed by Seehausen et al., 2014; Salzburger, 2018; Santos et al., 2023), the modes of speciation of other African fish groups such as *Enteromius*, have remained largely understudied. Based on the geographical distribution found in this study with a rather clear separation between lowland and highland species from the Lake Edward system, allopatric speciation seems the most parsimonious hypothesis. However, the old age of the split of the species from the Lake Edward system and neighbouring regions that we found based on Molecular Clock analysis (i.e. origin in Pliocene to early-Pleistocene) is not in accordance with the moment that the uplift in Uganda affected the hydrology of the systems of Lakes Edward and Victoria (i.e. 0.78 to 0.13 Mya, mid-Pleistocene) (Beadle 1974). Several hypotheses can be invoked to explain these estimated ages. Possibly, the evolutionary rate of *Enteromius* is much higher compared to other tropical fishes, and hence our age estimates are too old. Not much is known about the rate of evolution in *Enteromius*. Two other hypotheses are that the allopatric speciation occurred during an event much older then the change of hydrology due to rifting in the mid-Pleistocene and, lastly, that the highland specimens used to be isolated from the lowland specimens within the Lake Edward system and from the populations of *Enteromius* within the Lake Victoria system. Due to the uplifting and changing hydrology the highland specimens are now found in the two systems. To verify the last two hypotheses, a more profound biogeographical study will be performed.

The main problem in testing the hypotheses mentioned above is the lack of fossils of *Enteromius*. Fossils of '*Barbus*'-like pharyngeal teeth with an age of 12 to 13 million years were found in Africa. However, the identification of the fossils is complicated because only the pharyngeal teeth were found (Van Couvering, 1977; Yang et al., 2015). We used the divergence date between the small African diploid barbs and the Asian *Puntigrus tetrazona*, calculated in Ren & Mayden (2016) as 30.1 Mya, as a basis to put the lineages in a timeframe but the timing of the divergence between the African and Asian diploid barbs seems to be problematic. A recent study with fossil-based calibration points of polyploid cyprinids, found that species of *Enteromius* and *P. tetrazona* diverged 61.7 Mya (Yang et al., 2022), which is twice as long ago as suggested by Ren & Mayden (2016). Without a better fossil record

of *Enteromius*, it will remain difficult to place the speciation events within *Enteromius* in a geological timeframe and hypothesize their modes of speciation.

4.3. Sawfin barbs from a broader region

Our data shed a new light on the distribution patterns of the species of *Enteromius* from the Lake Edward and adjacent systems. With the inclusion of specimens (from GenBank) from other areas in the East Coast ichthyofaunal province and the basins of the Congo and the downstream part of the Nile, we also gained information on the relationship of the East African small barbs. However, understanding the relationship within *Enteromius* is complicated by the high number of misidentifications in online datasets (Hayes & Armbruster, 2017). In our study, we also highlighted several misidentifications and gave plausible explanations when possible (Table S4). Many species, which were misidentified, belonged to so-called wastebasket taxa (i.e. the assignment of specimens with a similar morphology to a certain species without further examination) (Decru et al., 2016). *Enteromius miolepis* and *E. paludinosus* are examples of such wastebasket taxa. In addition, species such as *E. paludinosus* have a very wide distribution pattern: from Ethiopia to South-Africa (Lévêque & Daget, 1984), which is in high contrast to the species which are endemic to certain stretches of a river as found in Van Ginneken et al. (2017). Based on our results, it is difficult to make any conclusions on the taxonomic position of the species of *Enteromius* from the Lake Edward system and neighbouring regions. However, we can propose one hypothesis. Based on the Cyt b gene, we found that specimens from the type locality of *E. neumayeri* are clustering together with *E.* sp. 'pellegrini Kafu' and *E.* sp. 'pellegrini lowland'. *Enteromius pellegrini* and *E. neumayeri* are morphologically similar. However the former is described from Lake Kivu, while the latter is described from an affluent of the Ewaso Ng'iro River (Kenya). We hypothesise that either *E.* sp. 'pellegrini Kafu' or *E.* sp. 'pellegrini lowland' might be conspecific with *E. neumayeri*. The distance between the type locality of *E. neumayeri* and the sample localities of *E.* sp. 'pellegrini lowland' and *E.* sp. 'pellegrini Kafu' is rather large. However, a similar pattern was observed by Maetens et al. (2020) where *E. cercops* (Whitehead, 1960), described from Luambwa, Nzoia River, Nyanza Province (Kenya) was synonymized with *E. alberti*, described from the Rutshuru River, May-Ya-Moto, Lake Edward Basin (DRC).

4.4. Conclusions

The results of this study underscore the recent findings of unrecognized diversity within *Enteromius*. Additionally, the discrepancy between the current geographical distribution patterns and the revealed

age of the species based on our molecular clock analysis prompts inquiries into how these species, despite their age, occur in such close proximity. In contrast with the study of the Kenyan small barbs by Schmidt et al. (2017), no introgression patterns were found. However, genome scale analyses could offer further insights. Preliminary genomic results revealed for the aberrant specimen (HP2674), which clusters genetically (based on COI) in *E.* sp. 'pellegrini highland', a position within *E.* sp. 'pellegrini lowland', meaning that a possible event of hybridisation took place. It is evident that *Enteromius* holds the potential to emerge as a new model organism for investigating speciation patterns in tropical riverine freshwater fishes.

References

Bamba M., Vreven E.J. & Snoeks J. (2011). Description of *Barbus teugelsi* sp. nov. (Cypriniformes: Cyprinidae) from the Little Scarcies basin in Guinea, Africa. *Zootaxa*, 2998: 48-65. https://org/10.11646/zootaxa.2998.1.4

Beadle L.C. (1974). *The Inland Waters of Tropical Africa: An Introduction to Tropical Limnology*. Longman Group LTD, London: 365 pp.

Bookstein F., Chernoff B., Elder R., Humphries J., Smith G. & Strauss R. (1985). Morphometrics in Evolutionary Biology: the geometry of size and shape change, with examples from fishes. *Journal of the Academy of Natural Sciences of Philadelphia*, Special publication, 15: 1-277.

Borchsenius F. (2009). FastGap 1.2

Brown W. M., George M. Jr., Wilson A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 76(4): 1967-71.

Chakona A., Malherbe W.S., Gouws G. & Swarts E.R. (2015). Deep genetic divergence between geographically isolated populations of the goldie barb (*Barbus pallidus*) in South Africa: potential taxonomic and conservation implications. *African Zoology* 50(1): 5-10.

Chernomor O., von Haeseler A. & Minh B.Q. (2016) Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, 65: 997-1008. https://doi.org/10.1093/sysbio/syw037

Clement M., Posada D. & Crandall K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, 9: 1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

Darriba D, Taboada GL, Doallo R, Posada D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9(8), 772.

Decru E., Moelants T., De Gelas K., Vreven E., Verheyen E. & Snoeks J. (2016). Taxonomic challenges in freshwater fishes: a mismatch between morphology and DNA barcoding in fish of the north-eastern part of the Congo Basin. *Molecular Ecology Resources*, 16: 342-352. https://doi.org/10.1111/1755-0998.12445

632 Decru E., Vranken N., Bragança P.H.N., Snoeks J. & Van Steenberge M. (2019). Where ichthyofaunal
633 provinces meet: the fish fauna of the Lake Edward system, East Africa. *Journal of Fish Biology*, 96(5):
634 1186-1201. https://doi.org/10.1111/jfb.13992

635 Decru E., Vranken N., Maetens H., Mertens De Vry A., Kayenbergh A., Snoeks J. & Van Steenberge M.
636 (2022). DNA barcoding the Lake Edward basin: high taxonomic coverage of a tropical freshwater
637 ichthyofauna. *Hydrobiologia*, 849: 1743-1762. https://doi.org/10.1007/s10750-022-04812-0.

638 Doornkamp, J. C., & Temple, P. H. (1966). Surface, drainage and tectonic instability in part of
639 southern Uganda. *The Geographical Journal*, 132 (2), 238–252.

640 Drummond A.J. & Rambaut A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC*
641 *Evolutionary Biology* 7(214). https://doi.org/10.1186/1471-2148-7-214

642 Drummond A.J., Suchard M.A., Xie D., Rambaut A. (2012) Bayesian phylogenetics with BEAUti and the
643 BEAST 1.7. *Molecular Biology and Evolution* 29, 1969-1973. DOI:10.1093/molbev/mss075

644 Edgar R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput.
645 *Nucleic Acid Research* 32(5): 1792-7. https://doi.org/10.1093/nar/gkh340

646 Englmaier G.K., Tesfaye G. & Bogutskaya N.G. (2020). A new species of *Enteromius* (Actinopterygii,
647 Cyprinidae, Smiliogastrinae) from the Awash River, Ethiopia, and the re-establishment of *E.*
648 *akakianus*. *Zookeys* 902: 107-150. https://doi.org/10.3897/zookeys.902.39606

649 Froese, R. and D. Pauly. Editors. (2024). FishBase. World Wide Web electronic publication.
650 www.fishbase.org, version (02/2024)

651 Greenwood, P. H. (1973). A revision of the *Haplochromis* and related species (Pisces: Cichlidae) from
652 Lake George, Uganda. Bulletin of the British Museum (Natural History) *Zoology*, 25, 139–242.

653 Greenwood P.H. (1983). The zoogeography of African freshwater fishes: bioaccountancy or
654 biogeography?. In: Sims et al. (eds) *Evolution, Time and Space: the Emergence of the Biosphere,*
655 *Systematics Association*, Special volume 23: 179-199. New York & London: Academic Press.

656 Hayes M.M. & Armbruster J.W. (2017). The taxonomy and relationships of the African small barbs
657 (Cypriniformes: Cyprinidae). *Copeia*, 105(2): 348-362. https://doi.org/10.1643/CI-15-348

658 Hebert P.D.N., Cywinska A., Ball S.L. & deWaard J.R. (2003). Biological identifications through DNA
659 barcodes. *Proceedings of the Royal Society of London B*. Biological Sciences 270: 313–321.
660 https://doi.org/10.1098/rspb.2002.2218

661 Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q. & Vinh L.S. (2018). UFBoot2: Improving the
662 ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35: 518–522.
663 https://doi.org/10.1093/molbev/msx281

664 Ivanova N.V., Zemlak T.S., Hanner R.H. & Hebert P.D. (2007). Universal primer cocktails for fish DNA
665 barcoding. *Molecular Ecology Notes*, 7: 544-548. https://doi.org/10.1111/j.1471-8286.2007.01748.x

666 Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A. & Jermiin L.S. (2017). ModelFinder: Fast
667 model selection for accurate phylogenetic estimates. *Nature Methods*, 14: 587–589.
668 https://doi.org/10.1038/nmeth.4285

669 Kambikambi M.J., Kadye W.T. & Chakona A. (2021). Allopatric differentiation in the *Enteromius*
670 *anoplus* complex in South Africa, with the revalidation of *Enteromius cernuus* and *Enteromius*

671 *oraniensis*, and description of a new species, *Enteromius mandelai* (Teleostei: Cyprinidae). *Journal of*
672 *Fish Biology* 99(3): 931-954. https://doi.org/10.1111/jfb.14780

673 Kasangaki A., Babaasa D., Efitre J., McNeilage A. & Bitariho R. (2006). Links between anthropogenic
674 perturbations and benthic macroinvertebrate assemblages in Afromontane forest streams in Uganda.
675 *Hydrobiologia* 563, 231-245. https://doi.org/10.1007/s10750-005-0009-8

676 Kassambara A. & Mundt F. (2020). Factoextra: Extract and Visualize the Results of Multivariate Data
677 Analyses. R package version 1.0.7.

678 Kisekelwa T., Snoeks J., Decru E., Schedel F.B.D., Isumbisho M. & Vreven E. (2022). A mismatch
679 between morphological and molecular data in lineages of *Enteromius* (Cypriniformes: Cyprinidae)
680 from the Lowa basin (East Democratic Republic of the Congo: DRC) with the description of a new
681 species. *Systematics and Biodiversity*, 20(1): 1-22. https://doi.org/10.1080/14772000.2022.2135630

682 Kumar S., Stecher G. & Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version
683 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33: 1870-1874.
684 https://doi.org/10.1093/molbev/msw054

685 Lavoué S., Miya M., Arnegard M.E., McIntyre P.B., Mamonekene V. & Nishida M. (2011). Remarkable
686 morphological stasis in an extant vertebrate despite tens of millions of years of divergence.
687 *Proceedings of the Royal Society B* 278, 1003-1008. https://doi.org/10.1098/rspb.2010.1639

688 Lavoué S. (2020). Origins of Afrotropical freshwater fishes. *Zoological Journal of the Linnean Society*
689 188(2), 345-411. https://doi.org/10.1093/zoolinnean/zlz039

690 Leigh J. W. & Bryant D. (2015). PopART: Full-feature software for haplotype network construction.
691 *Methods in Ecology and Evolution*, 6, 1110–1116. https://doi.org/10.1111/2041-210x.12410

692 Lévêque C. & Daget J. (1984). Cyprinidae. p. 217-342. In: Daget J. Gosse J.-P. & Thys van den
693 Audenaerde D.F.E. (eds.). *Check-list of the freshwater fishes of Africa (CLOFFA)*. ORSTOM, Paris and
694 MRAC, Tervuren. Vol. 1.

695 López J.A., Chen W.-J. & Ortí G. (2004). Esociform Phylogeny. *Copeia*, 3: 449-464.
696 https://doi.org/10.1643/CG-03-087R1

697 Maetens H., Van Steenberge M., Snoeks J. & Decru E. (2020). Revalidation of *Enteromius alberti* and
698 presence of *Enteromius* cf. *mimus* (Cypriniformes: Cyprinidae) in the Lake Edward system, East Africa.
699 *European Journal of Taxonomy*, 700: 1-28. https://doi.org/10.5852/ejt.2020.700

700 Mayden R.L., Chen W.-J., Bart H.L., Doosey M.H., Simons A.M., Tang K.L., Wood R.M., Agnew M.K.,
701 Yang L., Hirt M.V., Clements M.D., Saitoh K., Sado T., Miya M. & Nishida M. (2009). Reconstructing
702 the phylogenetic relationships of the earth's most diverse clade of freshwater fishes—order
703 Cypriniformes (Actinopterygii: Ostariophysi): A case study using multiple nuclear loci and the
704 mitochondrial genome. *Molecular Phylogenetics and Evolution*, 51: 500-514.
705 https://doi.org/10.1016/j.ympev.2008.12.015

706 Mbalassa, M., Nshombo, M., Kateyo, M. E., Chapman, L., Efitre, J., & Bwanika, G. (2015).
707 Identification of migratory and spawning habitats of *Clarias gariepinus* (Burchell, 1822) in Lake
708 Edward-Ishasha River watershed, Albertine Rift Valley, East Africa. *International Journal of Fisheries*
709 *and Aquatic Studies*, 2(3), 128–138.

McGee M.D., Borstein S.R., Meier J.I., Marques D.A., Mwaiko S., Taabu A., Kishe M.A., O'Meara B., Bruggmann R., Excoffier L. & Seehausen O.(2020). The ecological and genomic basis of explosive adaptive radiation. *Nature* 586, 75–79. https://doi.org/10.1038/s41586-020-2652-7

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A. & Lanfear R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*,37(5): 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mukasa-Tebandeke I.Z., Karume I., Ssebuwufu J., Wasajja H.Z., Nankinga R. & Habimana M. (2020). How variations in concentrations of metal ions and suspended solids downstream river Rwabakazi in Uganda can be used to study pollution. *Journal of Advances in Chemistry*, 17: 44-63. https://doi.org/10.24297/jac.v17i.8767

Odada E.O., Olago D.O. & Ochola W.O. (eds.) (2006). Environment for Development: An Ecosystems Assessment of Lake Victoria Basin Environmental and Socio-economic Status, *Trends and Human Vulnerabilities*. UNEP/PASS, Kenya, 192 pp.

Oksanen J., Blanchet F.G., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin P.R., O'Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Szoecs E. & Wagner H. (2019). vegan: Community Ecology Package. R package version 2.5-6.

Onyutha C., Amollo C.J., Nyende J., & Nakagiri A. (2021). Suitability of averaged outputs from multiple rainfall-runoff models for hydrological extremes: a case of River Kafu catchment in East Africa. *International Journal of Energy and Water Resources*.

Paradis E. & Schliep K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35 (3): 526-528. https://doi.org/10.1093/bioinformatics/bty633

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rambaut A, Drummond AJ, Xie D, Baele G and Suchard MA (2018) Posterior summarisation

Rambout A. (2018). FigTree. Tree Figure Drawing Tool Version 1.4.4. http://tree.bio.ed.ac.uk/

Reardon E. E. & Chapman L. J. (2009). Hypoxia and life-history traits in a eurytopic African cichlid. *Journal of Fish Biology*, 75(7): 1795-1815. https://doi.org/10.1111/j.1095-8649.2009.02429.x

Ren Q. & Mayden R.L. (2016). Molecular phylogeny and biogeography of African diploid barbs, '*Barbus'*, and allies in Africa and Asia (Teleostei: Cypriniformes). *Zoologica Scripta*, 45(6): 642-649. https://doi.org/10.1111/zsc.12177

Roberts T.R. (1975). Geographical distribution of African freshwater fishes. *Zoological Journal of the Linnean Society*, 57: 249-319. https://doi.org/10.1111/j.1096-3642.1975.tb01893.x

RStudio Team. (2016). RStudio: Integrated Development for R. RStudio Inc., Boston, MA. Available from http://www.rstudio.com/

Salzburger W. (2018). Understanding explosive diversification through cichlid fish genomics. *Nature Reviews Genetics* 19: 705-717. https://doi.org/10.1038/s41576-018-0043-9

Santos E., Lopes J. F. & Kratochwil C. F. (2023). East African cichlid fishes. *EvoDevo* 14(1). https://doi.org/10.1186/s13227-022-00205-5

748 Schedel F.D.B., Musilova Z., Indermaur A., Bitja-Nyom A., Salzburger W. & Schliewen U.K. (2022).
749 Towards the phylogenetic placement of the enigmatic African genus *Prolabeops* Schultz, 1941.
750 *Journal of Fish Biology* 101(5), 1333-1342. https://doi.org/10.1111/jfb.15205

751 Schmidt R.C., Bart Jr. H.L. & Nyingi W.D. (2017). Multi-locus phylogeny reveals instances of
752 mitochondrial introgression and unrecognized diversity in Kenyan barbs (Cyprininae: Smiliogastrini).
753 *Molecular Phylogenetics and Evolution*, 111: 35-43. https://doi.org/10.1016/j.ympev.2017.03.015

754 Schmidt R.C., Bart Jr. H.L. & Nyingi W.D. (2018). Integrative taxonomy of the red-finned barb,
755 *Enteromius apleurogramma* (Cyprininae: Smiliogastrini) from Kenya, supports recognition of E.
756 *amboseli* as a valid species. *Zootaxa*, 4482(3): 566-578. https://doi.org/10.11646/zootaxa.4482.3.8

757 Schmidt R.C., Dillon M.N., Kuhn N.M., Bart Jr. H.L. & Pezold F. (2019). Unrecognized and imperilled
758 diversity in an endemic barb (Smiliogastrini, *Enteromius*) from the Fouta Djallon highlands. *Zoologica*
759 *Scripta*, 48(5): 605-613. https://doi.org/10.1111/zsc.12362

760 Schraml E. & Tichy H. (2010). A new species of *Haplochromis*, *Haplochromis katonga* n. sp.
761 (Perciformes: Cichlidae) from the Katonga River, Uganda. *Aqua: International Journal of Ichthyology*
762 16(3): 81-92.

763 Seehausen O., Lippitsch E., Bouton N. & Zwennes H. (1998). Mbipi, the rock-dwelling cichlids of Lake
764 Victoria: description of three new genera and fifteen new species (Teleostei). *Ichthyological*
765 *Exploration of Freshwaters* 9: 129-228.

766 Seehausen O., Butlin R. K., Keller I., Wagner C. E., Boughman J. W., Hohenlohe P. A., Peichel C. L.,
767 Saetre G. P., Bank C., Brännström Å., Brelsford A., Clarkson C. S., Eroukhmanoff F., Feder J. L., Fischer
768 M. C., Foote A. D., Franchini P., Jiggins C. D., Jones F. C., … & Widmer A. (2014). Genomics and the
769 origin of species. *Nature Reviews Genetics* 15(3), 176–192. https://doi.org/10.1038/nrg3644

770 Simmons M.P. & Ochoterena H. (2000). Gaps as characters in sequence-based phylogenetic analysis.
771 *Systematic Biology*, 49(2): 369-381. https://doi.org/10.1093/sysbio/49.2.369

772 Skelton P.H., Tweddle D. & Jackson P.B.N. (1991). Cyprinids of Africa. In: Winfield I.J. & Nelson J.S.
773 (eds) *Cyprinid Fishes: Systematics, Biology and Exploitation*: 211–239. Chapman & Hall, London, UK.

774 Snoeks J., De Vos L. & Van den Audenaerde D.T. (1997). The ichthyogeography of Lake Kivu. *South*
775 *African Journal of Science*, 93(11): 579-584.

776 Stewart K.M. (2009). Fossil Fish from the Nile River and Its Southern Basins. In Dumont H.J. (editor),
777 *The Nile: Origin, Environments, Limnology and Human Use*. Springer: 677-704.

778 Sudasinghe H., Ranasinghe T., Herath J., Wijesooriya K., Pethiyagoda R., Rüber L. & Meegaskumbura
779 M. (2021). Molecular phylogeny and phylogeography of the freshwater-fish genus *Pethia* (Teleostei:
780 Cyprinidae) in Sri Lanka. *BMC Ecol Evo* 21 (203). https://doi.org/10.1186/s12862-021-01923-5

781 Swofford D.L. (2003). PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version
782 4. Sinauer Associates, Sunderland.

783 Tan M. & Armbruster J.W. (2018). Phylogenetic classification of extant genera of fishes of the order
784 Cypriniformes (Teleostei: Ostariophysi). *Zootaxa* 4476(1): 006-039.
785 https://doi.org/10.11646/zootaxa.4476.1.4

786 Van Couvering J.A.H. (1977). Early Records of Freshwater Fishes in Africa. *Copeia* 1: 163-166.
787 https://doi.org/10.2307/1443521

Formatted: English (United Kingdom)

788 Van Ginneken M., Decru E., Verheyen E. & Snoeks J. (2017). Morphometry and DNA barcoding reveal
789 cryptic diversity in the genus *Enteromius* (Cypriniformes: Cyprinidae) from the Congo basin, Africa.
790 *European Journal of Taxonomy*, 310: 1–32. https://doi.org/10.5852/ejt.2017.310

791 Van Steenberge M.W., Vanhove M.P., Chocha Manda A., Larmuseau M.H., Swart B.L., Khang'Mate F.,
792 & Volckaert F.A. (2020). 2020. Unravelling the evolution of Africa's drainage basins through a
793 widespread freshwater fish, the African sharptooth catfish *Clarias gariepinus*. *Journal of*
794 *Biogeography* 47: 1739–1754. https://doi.org/10.1111/jbi.13858

795 Verheyen E., Salzburger E., Snoeks J. & Meyer A. (2003). Origin of the Superflock of Cichlid Fishes
796 from Lake Victoria, East Africa. *Science* 300(5617), 325-329. https://doi.org/10.1126/science.1080699

797 Ward R.D. (2012). FISH-BOL, A Case Study for DNA barcodes. In: Kress W.J. & Erickson D.L. (eds) *DNA*
798 *barcodes. Methods and Prococols*: 423-439. Human Press.

799 Ward R.D., Hanner R. & Hebert P.D.N. (2009). The campaign to DNA barcode all fishes, FISH-BOL.
800 *Journal of Fish Biology*, 74: 329.356. https://doi.org/10.1111/j.1095-8649.2008.02080.x

801 Xiao W., Zhang Y., Liu H. (2001). Molecular Systematics of Xenocyprinae (Teleostei: Cyprinidae):
802 Taxonomy, Biogeography, and Coevolution of a Special Group Restricted in East Asia. *Molecular*
803 *Phylogenetics and Evolution* 18(2); 163-173.

804 Yang L., Sado T., Hirt M.V., Pasco-Viel E., Arunachalam M., Li J., Wang X., Freyhof J., Saitoh K., Simons
805 A.M., Miya M., He S. & Mayden R.L. (2015). Phylogeny and polyploidy: resolving the classification of
806 cyprinine fishes (Teleostei: Cypriniformes). *Molecular Phylogenetics and Evolution*, 85: 97-116.
807 https://doi.org/10.1016/j.ympev.2015.01.014

808 Yang L., Naylor G.J.P. & Mayden R.L. (2022). Deciphering reticulate evolution of the largest group of
809 polyploid vertebrates, the subfamily cyprininae (Teleostei: Cypriniformes). *Molecular Phylogenetics*
810 *and Evolution* 166 (107323). https://doi.org/10.1016/j.ympev.2021.107323