MR. GONTRAN SONET (Orcid ID : 0000-0001-7310-9574)

Article type : Resource Article

DNA barcoding fishes from the Congo and the Lower Guinean provinces: assembling a reference library for poorly inventoried fauna

Running title

DNA barcoding fishes from Central Africa

Gontran Sonet*[1], Jos Snoeks[2,3], Zoltán T. Nagy[1], Emmanuel Vreven[2], Gert Boden[2], Floris C. Breman[2,4], Eva Decru[2], Mark Hanssens[2], Armel Ibala Zamba[5], Kurt Jordaens[2,6], Victor Mamonekene[5], Tobias Musschoot[2], Jeroen Van Houdt[7], Maarten Van Steenberge[1,2,3], Soleil Lunkayilakio Wamuini[8,9], Erik Verheyen[1,6]

* Corresponding author (gsonet@naturalsciences.be)

[1] Royal Belgian Institute of Natural Sciences, OD Taxonomy and Phylogeny - JEMU, Vautierstraat 29, B-1000 Brussels, Belgium.
[2] Royal Museum for Central Africa, Department of Biology – Vertebrates, Entomology, JEMU, Leuvensesteenweg 13, B-3080 Tervuren, Belgium.
[3] KU Leuven, Laboratory of Biodiversity and Evolutionary Genomics, Ch. de Bériotstraat 32, B-3000 Leuven, Belgium.
[4] Wageningen University and Research, Biosystematics Group, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.
[5] Université Marien Ngouabi, Institut de Développement Rural, B.P. 69 Brazzaville, Republic of the Congo.
[6] University of Antwerp, Department of Biology - Evolutionary Ecology Group, Universiteitsplein 1, B-2610 Antwerp, Belgium.
[7] UZLeuven, Genomics Core KULeuven, Gasthuisberg O&N1, 49 Herestraat - box 602, B-3000 Leuven, Belgium.
[8] I. S. P. Mbanza-Ngungu, Département de Biologie, B.P. 127, Mbanza-Ngungu, Democratic Republic of the Congo.
[9] University of Liège, Functional and Evolutionary Morphology Laboratory, Place du 20-Août 7, B-4000 Liège, Belgium.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1755-0998.12983

Abstract

The Congolese and Lower Guinean ichthyological provinces are understudied hotspots of the global fish diversity. Here, we barcoded 741 specimens from the Lower and Middle Congo River and from three major drainage basins of the Lower Guinean ichthyological province, Kouilou-Niari, Nyanga and Ogowe. We identified 194 morphospecies belonging to 82 genera and 25 families. Most morphospecies (92.8%) corresponded to distinct clusters of DNA barcodes. Of the four morphospecies present in both neighbouring ichthyological provinces, only one showed DNA barcode divergence <2.5%. A small fraction of the fishes barcoded here (12.9% of the morphospecies and 16.1% of the barcode clusters representing putative species) were also barcoded in a previous large-scale DNA analysis of freshwater fishes of the Lower Congo published in 2011 (191 specimens, 102 morphospecies). We compared species assignments before and after taxonomic updates and across studies performed by independent research teams and observed that most cases of inconsistent species assignments were due to unknown diversity (undescribed species and unknown intraspecific variation). Our results report more than 17 putative new species and show that DNA barcode data provide a measure of genetic variability that facilitates the inventory of underexplored ichthyofaunae. However, taxonomic scrutiny, associated with revisions and new species descriptions, is indispensable to delimit species and build a coherent reference library.

Keywords

Biodiversity, Freshwater, Taxonomy, COI, Ichthyofauna, Central Africa

1 Introduction

The Fish Barcode of Life initiative (FISH-BOL) is an international research collaboration aiming at assembling a standardized reference library of DNA barcodes for all fish species (Becker, Hanner, & Steinke, 2011; Hanner, Desalle, Ward, & Kolokotronis, 2011; Ward, Hanner, & Hebert, 2009). The goal of this campaign is to allow fish species identification through the comparison of query

sequences against the reference sequence database in the Barcode of Life Data Systems, BOLD (Ratnasingham & Hebert, 2007). DNA barcode libraries also facilitate species discovery when morphology alone is insufficient (April, Mayden, Hanner, & Bernatchez, 2011) and particularly in poorly inventoried areas characterized by taxonomically hyperdiverse faunas (L. H. Pereira, Hanner, Foresti, & Oliveira, 2013; Pugedo, de Andrade Neto, Pessali, Birindelli, & Carvalho, 2016).

One of the main knowledge gaps on global fish diversity concerns the two ichthyofaunal provinces that make up Central Africa: the Congo Basin and Lower Guinean provinces (Brooks, Allen, & Darwall, 2011; Darwall et al., 2011; Lévêque, Oberdorff, Paugy, Stiassny, & Tedesco, 2008; Roberts, 1975). The Congo Basin is the second largest catchment area in the world after the Amazon Basin and is characterized as a hotspot of fish diversity (Snoeks, Harrison, & Stiassny, 2011), with about 1000 described species from the region excluding lakes Kivu and Tanganyika and the Malagarazi system (Froese & Pauly, 2018). Despite more than a century of taxonomic efforts with numerous field expeditions, local inventories and new species descriptions (e.g. Boulenger, 1901; Decru et al., 2017; Roberts & Stewart, 1976; Shumway et al., 2003; Stiassny & Mamonekene, 2007; Van Steenberge, Vreven, & Snoeks, 2014; Wamuini, Vreven, Vandewalle, Mutambue, & Snoeks, 2010), its fauna remains poorly documented and large areas of the Congo Basin remain underexplored (Thieme et al., 2005). The Lower Guinean province has been more intensively studied in the last decades (e.g. Stiassny, Teugels, & Hopkins, 2007; Walsh & Mamonekene, 2014). Although some of its river basins show ichthyofaunal similarities with the Congo Basin (Brooks et al., 2011), its southern part is characterized by very high levels of endemism. The development of a DNA reference library for the ichthyofauna of Central Africa is a complex and slow process, which is fraught with difficulties. The main DNA barcoding studies focusing on this region established DNA barcoding libraries for 328 species of the Lower Congo (Lowenstein, Osmundson, Becker, Hanner, & Stiassny, 2011) and for 206 species of the north-eastern part of the Congo Basin (Decru et al., 2016). Both studies reported species delimitation issues, which are primarily due to limited exploration throughout the region, difficult compilation of identification keys impeding identification of

specimens to the species level, lack of species distribution data, and a chronic lack of up-to-date taxonomic and ecological knowledge. As a consequence, the DNA barcoding of fishes from these regions is exposed to a plethora of species identification problems which should be addressed with taxonomic approaches (e.g. Decru, Vreven, & Snoeks, 2012, 2013; Lowenstein et al., 2011; Vreven, Musschoot, Snoeks, & Schliewen, 2016).

In line with the recommendations of the African Regional Working Group of FISH-BOL (2008), we used DNA barcoding as a tool to inventory the fish biodiversity of several areas of the Congo Basin and Lower Guinean ichthyological provinces. Our team comprises experienced fish taxonomists that are deeply involved in the revision of the Afrotropical fish fauna. The combined use of morphological characters and DNA barcodes is recognized as an appropriate strategy to improve the reliability of species identifications (Hubert & Hanner, 2015; Janzen et al., 2009; Sheth & Thaker, 2017). We pursue two objectives: (1) to link DNA barcode sequences to morphospecies and enrich existing taxonomical data related to Congolese and Lower Guinean fish biodiversity; (2) to assess whether DNA barcode data confirm the presumed conspecificity of morphologically identified species within and across ichthyological provinces. We also assess the dynamics of building a DNA barcode library for freshwater fishes by (1) evaluating whether DNA barcode libraries that are generated independently for the same part of the Congolese drainage system (Lowenstein et al., 2011) can be combined to ascertain unambiguous species identifications and (2) comparing DNA barcoding results obtained for the same data set before (in 2012) and after (in 2018) recent taxonomic updates.

2 Materials and Methods

2.1 Sampling

A total of 741 specimens were sampled during four field campaigns carried out between 2004 and 2007 in the Democratic Republic of the Congo (DRC) and in the Republic of the Congo (Congo-Brazzaville) (Fig. 1 and Table S1). In the Congolese ichthyological province (CO), we sampled a total

of 496 specimens in the drainage basins of the Djoué (right bank, Congo-Brazzaville), Inkisi (left bank, DRC), Luki (right bank, DRC), and Léfini (right bank, Congo-Brazzaville). The three first tributaries represent portions of the Lower Congo whereas the Léfini is situated in the Middle Congo. Other sampling sites included sections of the mainstream of the Lower Congo and Pool Malebo, which belongs to the Middle Congo. In the Lower Guinean ichthyological province (LG), we sampled 245 specimens in three major drainage systems, viz. Kouilou-Niari, Nyanga and Ogowe (including Ogowe, Polo-Ogowe and Ngongo).

Fishes were caught using gill nets. A representative selection was individually tagged using nylon T-bar anchor tags (Hallprint, Australia) and used for DNA analysis. Fin clips were sampled and individually stored in absolute or highly concentrated ethanol. Voucher specimens were fixed in formalin and are deposited at the Royal Museum for Central Africa (RMCA, Tervuren, Belgium).

2.2 Morphological species identifications

All specimens were identified as morphospecies based on their external morphology, a task that proved to be rather difficult because of the lack of identification keys for some areas or groups, the presence of potential undescribed species and vague morphological species boundaries. This task was achieved using the available taxonomic literature listed in the Supporting Information of Decru et al. (2016) or through direct comparison with type specimens and other voucher specimens at the RMCA. Specimens that did not entirely correspond to a known species were indicated with ‘cf.’ or ‘sp.’ and were treated as different morphospecies in the DNA barcoding analyses (see below). We used ‘cf.’ for specimens resembling the nominal species but showing at least one diagnostic morphological character that deviated to the extent that we were unable to decide whether this represented geographical variation or pointed to a possible undescribed species. We used ‘sp.’ when

the examined specimens most likely represented a species new to science. We also used 'x' for one specimen that was considered as an hybrid: *Coptodon tholloni* x *rendalli*

2.3 DNA data collection

Total genomic DNA was extracted using DNeasy Blood & Tissue kits (Qiagen, The Netherlands) following standard protocols for animal tissue. DNA quantity and purity was checked using the ND-1000 spectrophotometer (NanoDrop Technologies, USA). To amplify the standard animal DNA barcode region—the 5'-end (Folmer, Black, Hoeh, Lutz, & Vrijenhoek, 1994) of the cytochrome *c* oxidase subunit I (COI) gene—as suggested by Hebert, Cywinska, Ball and DeWaard (2003), we used the tailed version of the primer pair originally published by Ward, Zemlak, Innes, Last and Hebert (2005), FishF1_t1 (TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC) and FishR2_t1 (CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA) using tails M13 forward (–21) and M13 reverse (–27) (Messing, Crea & Seeburg, 1981).

The PCR profile was as follows: 94 °C for 3 min; 35–40 cycles of 94 °C for 40 s, 53 °C for 40 s and 72 °C for 60 s; 72 °C for 7 min, and subsequent storage of the samples at 4 °C. PCR products were visualized using 1.2% agarose gel electrophoresis. Purification was done either on illustra GFX PCR DNA purification kit columns (GE Healthcare, USA) or using the NucleoFast 96 PCR Plate (Macherey-Nagel, Germany) and vacuum-purification. PCR products were sequenced bi-directionally using the M13 vector primers. DNA sequencing was carried out on ABI automated capillary sequencers using BigDye v1.1 or BigDye v3.1 chemistry following the manufacturer's instructions (Life Technologies, USA). DNA sequences were quality-checked, assembled and aligned with the SeqScape v2.5 software (Life Technologies, USA). Sequences obtained from type specimens were given the labels holo- and paragenetypes following Chakrabarty (2010).

2.4 DNA analyses

Analyses were performed on the COI data set produced here (data set A) and on a combined data set (data set A+B) obtained after merging our data with the records published by Lowenstein et al. (2011). This latter data set (B) represents the largest data set of freshwater fish sequences publicly available for Central Africa. It was assembled by an independent research team. It was compiled by selecting all samples from the American Museum of Natural History (AMNH) downloaded from GenBank using the keywords 'Barcode', 'Congo' and 'fish' (Table S1).

Using MEGA v6.06 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) and the R package ape (Paradis, Claude, & Strimmer, 2004), we calculated pairwise uncorrected p-distances (Srivathsan & Meier, 2012) and assessed the existence of a gap between distances within morphospecies and among morphospecies (Collins, Boykin, Cruickshank, & Armstrong, 2012). We then applied four methods to determine molecular operational taxonomic units (MOTUs) from the DNA barcodes. All work without any *a priori* knowledge of species identity and were developed to approximate putative species: the automatic barcode gap discovery (ABGD) method (Puillandre, Lambert, Brouillet, & Achaz, 2012), the Refined Single Linkage (RESL) analysis (Ratnasingham & Hebert, 2013), the General Mixed Yule Coalescent (GMYC) model (Pons et al., 2006) and a Bayesian implementation of the Poisson tree processes (bPTP) model (Zhang, Kapli, Pavlidis, & Stamatakis, 2013). ABGD, GMYC and bPTP were used to analyse data sets A and A+B whereas RESL was only used for data set A (tool currently not available for multiple selections). The ABGD method was used with default settings (prior maximal intraspecific distances between 0.001 and 0.1) and using uncorrected p-distances to automatically detect gaps in the distribution of pairwise distances among DNA barcodes, which can be used to delimit hypothetical species. The RESL analysis was performed using the BOLD System (Ratnasingham & Hebert, 2007). It first performs a single linkage clustering of the records and then clusters records with high sequence similarity and connectivity, and separates those with lower similarity and sparse connectivity. The GMYC (single threshold approach) and the bPTP models are

based on the phylogenetic species concept. The first identifies the transition points between inter- and intra-species branching rates on a time-calibrated ultrametric tree whereas the second is based on a transition in the number of substitutions and does not require an ultrametric tree. Bayesian inference (BI) of phylogeny and the maximum likelihood (ML) method were used to reconstruct the phylogenetic trees used for GMYC and bPTP, respectively. The first phylogenetic analysis was conducted using the Yule model and a constant clock in BEAST2 (Bouckaert et al., 2014). The second analysis was conducted using RAxML (Stamatakis, 2015) with 1000 bootstrap pseudo-replicates. Both analyses were run on the CIPRES Science Gateway (Miller, Pfeiffer, & Schwartz, 2010) using the best partition scheme and best-fit substitution models estimated using PartitionFinder v. 1.1 (Lanfear, Calcott, Kainer, Mayer & Stamatakis, 2014). For BI, two parallel runs were run for 20 million generations. Convergence was checked and the first 25% of the trees were discarded ("burn-in").We finally evaluated the agreement between the morphospecies and the DNA barcode clusters obtained above and those obtained from a tree-based cluster analysis. For the latter analysis, distance-based neighbour joining (NJ) trees were constructed using MEGA v6.06 (Tamura et al., 2013) for the data sets A and A+B and for the five families that are represented by the highest number of sequences in the data set (>60 sequences per family for Alestidae, Cichlidae, Cyprinidae, Distochodontidae and Mormyridae). Node support of the NJ trees was evaluated by non-parametric bootstrapping using 1000 replicates. Given the conceptual problems associated with the interpretation of NJ trees when delimiting species (Meier, Shiyang, Vaidya, & Ng, 2006), we did not use the NJ tree to delimit MOTUs. Rather, we used it to assess if all DNA barcodes obtained for each morphospecies clustered (with a bootstrap value > 95%) and were not mixed with DNA barcodes from other morphospecies in the tree. This analysis was performed for data sets A and A+B and using the taxonomical classifications available in 2012 and in 2018. Obviously, morphospecies that were represented by one single sequence (singleton) were not analysed as clusters. They were only considered if they were mixed with other sequences.

3 Results

3.1 The new DNA barcode library (data set A)

We obtained DNA barcodes of 589–652 base pairs (bp) for 741 specimens (Process ID from BCOVR001-17 to BCOVR741-17 in the Barcode of Life Data Systems, BOLD), representing 194 morphospecies, 82 genera, 25 families and 10 orders (Table S1). The number of sequences obtained per morphospecies averaged 3.82 and ranged from 1 to 27 (55 morphospecies were represented by singletons). Pairwise p-distances observed within each morphospecies (mean=0.96%) were generally much lower than those observed among morphospecies (20.7%) (Fig. 2A). However, there was no barcoding gap as their ranges overlapped (0–10.7% and 0–28.1%, respectively). For example, p-distances within *Clarias camerunensis* (Clariidae) (0–7.2%) exceeded the interspecific p-distances between *Marcusenius moorii* and *M. kutuensis* (Mormyridae) (3.4–4.6%). The numbers of MOTUs identified on the exclusive basis of COI were 207–210 (ABGD with prior maximal distances of 0.013–0.001), 207 (RESL), 204–212 (GMYC) and 186–220 (bPTP). All methods produced very similar partitions (Table S1). For example, in the partition considering 207 MOTUs, 17 morphospecies were split in different putative species, while 18 were lumped with other morphospecies. In the same partition, distances among barcodes of the same MOTUs (0–1.9%) overlapped with those of different MOTUs (1.5–28.1%), but to a lower extent than when comparisons were based on morphospecies (Fig. 2A). The clustering obtained in the NJ tree concurred with 180 of the 194 morphospecies (92.8%) analysed here (Table S2 and Fig. S1). The 14 remaining morphospecies (six with provisional names and eight representing nominal species) were split in different clusters, or lumped (and sometimes shared their DNA barcodes) with other morphospecies (see below).

When the same data set was analysed with the taxonomic classification available in 2012, the proportion of morphospecies in agreement with the clusters found in the NJ tree (176/197=89.3%) was lower than the results of 2018 (92.8%, see above).

In some cases, the morphospecies defined after the taxonomic update of 2018 were still not matching the clusters of the NJ tree. The catfish *Clarias angolensis* (Clariidae) contained at least two haplotype groups in the Lefini (divergence of 4.1–6.1% between the groups) but one group clustered with DNA barcodes of *Clarias gabonensis* from the Inkisi (2.4–2.6% divergence between *Clarias gabonensis* and its closest cluster *Clarias angolensis*). Four *Ctenopoma* (Anabantidae) species, *C. ocellatum*, *C.* cf. *maculatum* (singleton), *C. acutirostre* and *C.* sp. 'Lefini' (singleton) shared the same haplotype. Similarly, DNA barcodes of *Coptodon tholloni* (Cichlidae) clustered with *C. congicus* (0.2–0.5%). Besides, DNA barcodes for *Petrocephalus simus* (Mormyridae) from the Ogowe (LG) were more similar (2.6%) to *P. binotatus* (Lefini, CO, singleton) than to presumed conspecifics occurring in the Kouilou-Niari, LG (3.4–3.6%). In some cases, morphospecies showing minor morphological differences ('cf.' in identification) clustered or shared COI haplotypes: *Petrocephalus microphthalmus* (singleton, Mormyridae) shared a common haplotype with *Petrocephalus* cf. *microphthalmus* (singleton). This also held for *Malapterurus beninensis* (Malapteruridae) and *Malapterurus* cf. *beninensis* (singleton) and for *Labeobarbus* sp. 'intermediate' and *L.* sp. 'Inkisi' (Cyprinidae).

3.2 Morphospecies occurring in neighbouring ichthyological provinces (data set A)

Only four of the 194 morphospecies sequenced here were collected both in the Lower Guinean (LG) and in the Congo (CO) provinces (Fig. 3A). In one of them, *Marcusenius moorii* (Mormyridae), DNA barcode divergences among ichthyological provinces (0.7–1.7%) were in the range of the divergences within ichthyological provinces (0–1.9%) (Fig. 2B). In the three other morphospecies, specimens from different ichthyological provinces showed important DNA barcode divergences (up to 9.4%) although they did not show any conspicuous morphological differences. For *Clarias camerunensis* (Clariidae), the DNA barcode divergence between the specimens of CO and LG reached 7.2%. For *Hemichromis elongatus* (Cichlidae), the DNA barcodes obtained from specimens of LG were highly divergent from those of CO (6.0–9.4%). The populations of LG were divided in

three DNA barcode clusters (4.8–9.2% among clusters), not related to their geographical origin, while the CO populations from the Inkisi and the Lefini shared one haplotype (Fig. 3A). For *Mastacembelus niger* (Mastacembelidae), four clusters of DNA barcodes were observed (divergences of 0–0.5% within clusters and 1.9–5.5% among clusters), each cluster representing specimens collected in a different river: Kouilou-Niari (LG), Polo-Ogowe (LG), Inkisi (CO) and Lefini (CO) (Fig. 3A). Interestingly, divergences between rivers of the same ichthyological province (5.3–5.5%) were larger than those between rivers of adjacent provinces (1.9–4.6%).

In contrast, four undescribed morphospecies showed both morphological differences and DNA barcode divergences with the morphologically closest nominal species found in the neighbouring province (Fig. 3B): *Pareutropius* sp. '*debauwi*-like' (Schilbeidae, LG) showed divergences of 8.9–9.4% with *Pareutropius debauwi* (CO); *Coptodon* cf. *rendalli* (Cichlidae, LG) with *Coptodon rendalli* (CO) (2.9–3.1%); *Paramormyrops* cf. *kingsleyae* (Mormyridae, CO) with *Paramormyrops kingsleyae* (LG) (5.3–5.5%); *Garra* cf. *ornata* (Cyprinidae, CO) with *Garra ornata* (LG) (4.8%). For two other hitherto undescribed morphospecies that resembles species from the neighbouring province, *Labeo* cf. *lukulae* (LG) and *Chiloglanis* cf. *batesii* (CO), no DNA barcode was available for representatives of the neighbouring province.

3.3 Morphospecies from a single ichthyological province (data set A)

Our results also showed high diversity within a single ichthyological province (Fig. 4A). For example, two undescribed congeneric mormyrids with strikingly different morphologies and divergent DNA barcodes (9.9%) were found in the Kouilou-Niari (LG): *Ivindomyrus* sp. 'elongate' (singleton) and *Ivindomyrus* sp. 'short' (singleton). Also, one undescribed distichodontid was found in the Nyanga (LG): *Nannocharax* sp. 'Nyanga' (singleton, >7.2% divergence with the other *Nannocharax*). Seven other morphospecies with provisional names showed relatively high DNA barcode divergences with

the morphologically most similar formally described species (Fig. 4A). Six of them were found in different river basins: *Amphilius* cf. *nigricaudatus* (Amphiliidae, Nyanga, LG) showed a divergence of 8.7% with *Amphilius nigricaudatus* from Kouilou-Niari (LG). *Nannocharax* sp. '*parvus*-like' (Distichodontidae, Polo-Ogowe, LG, singleton) showed a divergence of 13.5% with *Nannocharax parvus* from Nyanga (LG, singleton). *Enteromius* sp. '*miolepis*-like' (Cyprinidae, Inkisi, CO) showed divergences of 4.6–4.8% with *Enteromius miolepis* from Lefini (CO). *Chilochromis* sp. '*dupontiOgowe*' (Cichlidae, Ogowe, LG), *C*. sp. '*dupontielongate*' (Kouilou-Niari, LG) and *C*. sp. '*dupontideep*' (LG, Nyanga) were found in three different river basins and showed DNA barcode divergences of 1.4–3.4%. The same was observed within a single river (Kouilou-Niari, LG) for a seventh morphospecies: *Parauchenoglanis* sp. '*balayi*-like' (Claroteidae) showing 10.6–10.8% divergence with *Parauchenoglanis balayi*.

Considerable DNA barcode divergences (0.7–3.8%) were also detected within nine morphospecies (Fig. 4B). Four of them were from different river basins of LG (Fig. 4B): *Atopochilus savorgnani* (Mochokidae, divergences of 0.5–1.4%), *Doumea* cf. *sanaga* (Amphiliidae, 0–2.6%), *Distichodus hypostomatus* (Distichodontidae, 0–1.2%) and *Opsaridium ubangiense* (Cyprinidae, 1.2–3.8%). In three species of CO, we detected different haplotype groups (each haplotype group represented by 2–4 haplotypes) within the single sub-drainage basin of the Lefini (Fig. 4C): *Epiplatys multifasciatus* (Nothobranchiidae, 1.7–2% divergence among the haplotype groups), *Gnathonemus petersii* (Mormyridae, 0.5–1.0%) and *Pollimyrus nigripinnis* (Mormyridae, 0.5–3.8%). In the two last morphospecies, divergences were observed within the same river (Fig. 4C): *Schilbe grenfelli* (Schilbeidae) from the Lefini river (CO) was represented by two haplotypes (1.4% divergence), one of which was also found in the Inkisi (CO), and *Chromidotilapia* cf. *kingsleyae* (Cichlidae, LG) was represented by two well-differentiated clades found in sympatry in the Ogowe and in the Nyanga (0–0.7% and 2.9–3.4% divergence within and between the clades, respectively).

In contrast, we also reported small COI distances between *Parauchenoglanis* sp. '*balayi*-like' (Claroteidae) and *P.* cf. *pantherinus* (singleton) (0.7-1%, Fig. 4A). More surprisingly, the intergeneric divergence found between *Oreochromis niloticus* (nine DNA barcodes) and *Sarotherodon galilaeus* (singleton) (both Cichlidae) was very small (0.7%) compared to the interspecific divergence found within *Oreochromis* (9.2% between *O. niloticus* and *O. schwebischi*).

3.4 The combined DNA barcode library (data set A+B)

Our data set was merged with the publicly available data set of Lowenstein et al. (2011) containing 191 DNA barcodes of freshwater fishes from the Lower Congo (data set B). This data set represented 102 morphospecies, 48 genera (after updating the taxonomy – Table S1), 18 families and 8 orders. The number of sequences available per morphospecies averaged 1.87 and ranged from one to eight, with 47 morphospecies represented by a single sequence. The combined DNA barcode library (data set A+B) counted 932 DNA sequences representing 263 morphospecies, 96 genera, 27 families and 12 orders (average of 3.51 specimens sequenced per species, ranging from 1 to 27). In the combined data set there were 81 singletons, which was a lower number than the sum of the singletons in both data sets (56 for data set A and 47 for B, respectively). The number of morphospecies that were present in both data sets was 34 (12.9% of all morphospecies).

Barcode distances within and among morphospecies (0–23.4 and 0–28.7%, respectively) overlapped even more than in data set A (Fig. 2A). The numbers of MOTUs identified on the exclusive basis of COI were 261–266 (ABGD with prior maximal distances of 0.013–0.001), 253–265 (GMYC) and 229–267 (bPTP). t. In the partition considering 261 MOTUs, distances among barcodes of the same MOTUs (0–2.4%) overlapped with those of different MOTUs (1.2–28.7%), but to a lower extent than when comparisons were based on morphospecies (Fig. 2A). In the same partition, 42 MOTUs (16.1% of all MOTUs) were represented both in A and B. We also observed that 19 morphospecies were split

in two or more putative species, 61 morphospecies were lumped with at least one other morphospecies and 16 morphospecies were both split and lumped with other putative species (Table S2).

In the NJ tree (Table S2 and Fig. S2), the proportion of morphospecies clustering in distinct clusters decreased to 78.3% (92.8% in data set A and 78.4% in data set B). We observed that 21 of the 57 morphospecies that were lumped or split in the combined NJ tree were not problematic in the separate data sets (Table S2))). They belonged to the Anabantidae (2 morphospecies), Clariidae (1), Cyprinidae (8), Distichodontidae (1), Mormyridae (7) and Citharinidae (2). Some of these new inconsistencies were caused by a lumping of morphospecies with provisional names ('sp.' or 'cf.'): *Labeobarbus* sp. 'intermediate' and *L.* sp. 'Inkisi'(A) clustered with *L. stenostoma* (B). *Garra* cf. *ornata* (A) was lumped with *G. ornata* (B), *Ctenopoma* cf. *nigropannosum* (A) with *C. gabonense* (B), and *Pollimyrus* cf. *nigripinnis* (A) with *P. maculipinnis* (B). Other inconsistencies were due to the lumping of nominal species in the NJ tree: *Labeo lineatus* (B) with *L. greenii* (singleton, A), *Mormyrops furcidens* (A) with *M. lineolatus* (B)*, M. masuianus* (B) with *M. sirenoides* (A)*, Pollimyrus nigripinnis* (A) with one haplotype of *P. maculipinnis* (B) and *Citharinus gibbosus* (B) with *C. macrolepis* (A). In another case, *Enteromius holotaenia* (B) was lumped with *E. miolepis* (A), which was split from the *E. miolepis* (B) identified independently. Similarly, *Clarias gabonensis* (singleton, B) was split from *C. gabonensis* (A) but lumped with some haplotypes of *C. angolensis* (A) identified here. Finally, three species were split in different clusters (*Enteromius rubrostigma*, *Opsaridium ubangiense* and *Eugnathichthys macroterolepis*).

Below, we limit ourselves to the comparison of morphospecies assignments and DNA barcodes for families for which we amassed the highest number of sequences: the Alestidae (73 sequences), Cichlidae (126), Cyprinidae (215), Distichodontidae (130) and Mormyridae (102). Alestidae: The DNA barcodes of the only morphospecies present in both data sets (*Phenacogrammus interruptus*) matched (Fig. 5). Cichlidae: Both data sets contained *Coptodon tholloni* and *C. congicus*. Although

some DNA barcodes from the two species were identical in our study, they were distinct (15.2%) in the data set B. Hence, the four DNA barcodes (3 *C. tholloni* and 1 *C. congicus*) of Lowenstein et al. (2011) appeared to allow the identification of these two species while these taxa could not be identified using our larger sampling (10 *C. tholloni* and 10 *C. congicus*, Fig. 6). Cyprinidae: Both data sets had six morphospecies in common. The DNA barcodes of three of these (*Enteromius rubrostigma*, *E. miolepis* and *Opsaridium ubangiense*) were split in different clusters after combining the two data sets. The three others morphospecies (*Labeo altivelis*, *Labeobarbus macrolepidotus* and *Raiamas kheeli*) clustered in the NJ tree in accordance with the morphospecies assignments). Five other morphospecies (*Enteromius holotaenia* (B), *Garra* cf. *ornata*, (A), *Labeo greenii* (A), *L. lineatus* (B) and *Labeobarbus stenostoma* (B) were lumped with other morphospecies in the combined data set. Two morphospecies associated to one valid species name (*Labeo annectens* and *L.* cf. *annectens*) were collected in different ichthyological provinces (each team in a different province) and showed DNA barcodes with divergences of 4.6–5.5%. (Fig. 7). Distichodontidae: Seven out of the eight species represented in both data sets clustered according to their morphospecies identification. Only *Eugnathichthys macroterolepis* was represented by DNA barcodes that were split and showed divergences of 10.84% (Fig. 8). Mormyridae: All seven morphospecies found in both datasets clustered according to their morphospecies identification. However, other morphospecies were lumped in the NJ tree: *Pollimyrus maculipinnis* (B) was lumped with with *P. nigripinnis* (A), *Mormyrops masuianus* (B) with *M. sirenoides* (A) and *M. lineolatus* (B) with *M. furcidens* (A) (Fig. 9).

4 Discussion

4.1 Undescribed diversity in the new DNA barcode library (data set A)

With this DNA barcode library of 741 fishes collected from the Lower and Middle Congo River (CO) and three major drainage basins of the Lower Guinean (LG) ichthyological province, we associate 194 morphospecies with DNA barcodes. Our NJ tree analysis showed that most morphospecies (92.8%) were resolved as distinct clusters of DNA barcodes. However, taxonomic assignment

remains a major challenge in this understudied species-rich fauna. Indeed, our analyses (ABGD, RESL, GMYC, bPTP and NJ tree) revealed some inconsistencies between morphology-based identifications and DNA barcode clustering. A considerable proportion of these inconsistencies can be attributed to undescribed diversity and, more specifically, to species that are either not yet described (splitting in the NJ tree) or to species whose phenotypic variation is not yet known (lumping in the NJ tree). Indeed, the results obtained for the same data set but based on the taxonomic knowledge of 2012 provided a larger number of inconsistencies than those based on the current taxonomy. This is mainly due to the revision of *Hepsetus* (Decru et al., 2013; Decru, Snoeks, & Vreven, 2015), *Congolapia bilineata* (Dunz, Vreven, & Schliewen, 2012) and the synonymization of *Varicorhinus* and *Labeobarbus* (Berrebi, Chenuil, Kotlík, Machordom, & Tsigenopoulos, 2014; Tsigenopoulos, Kasapidis, & Berrebi, 2010; Vreven et al., 2016). Other cases of overlooked biodiversity were suggested by our NJ tree and were already reported by Decru et al. (2016). For example, in the genus *Clarias*, specimens identified as *C. angolensis* represent three different lineages belonging to a cluster also including *C. gabonensis*.

The DNA barcodes obtained here suggest the existence of more than 17 undescribed species. Most of them (14) were already revealed by morphological analysis and supported as separated MOTUs in the DNA barcoding analyses: three collected in a single location and not associated to any other known species, four diverging from a species living in the neighbouring ichthyological province and seven diverging from a species living in another river of the same province (DNA barcodes showing 1.4–13.5% divergences). Three additional undescribed species are suggested by DNA barcode divergences >5% within three of the four morphospecies present in both ichthyological provinces. Other overlooked taxa may be distinguished by DNA barcode divergences of 0.7–3.8% within nine morphospecies present in a single province. The contribution of DNA barcoding for the detection of new taxa was acknowledged in most DNA barcoding campaigns focusing on large river basins (Decru et al., 2016; Hubert et al., 2008; Nascimento et al., 2016; Pereira, Pazian, Hanner, Foresti, & Oliveira, 2011; Shen, Guan, Wang, & Gan, 2016). Multiple lineages were suspected within *Enteromius*

miolepis on the basis of DNA barcodes. They could subsequently be distinguished using morphometric analyses and may represent undescribed species (Van Ginneken, Decru, Verheyen, & Snoeks, 2017). Conversely, phenotypic differences between *Labeobarbus* sp. 'Inkisi' and *L.* sp. 'intermediate' or *Coptodon congicus* and *C. tholloni* were not associated with DNA barcode divergences. Such cases have to be investigated with additional sampling and alternative DNA markers in order to check if they represent species that cannot be identified using DNA barcodes or species with large intraspecific phenotypic variation as reported in *Labeo altivelis* (Van Steenberge, Gajdzik, Chilala, Snoeks, & Vreven, 2017). Even if most of these taxonomic investigations will result in a better agreement between species and MOTUs (suggesting that the ranges of intra- and interspecific barcode distances would become more similar to those observed for MOTUs in Fig. 2A), several species delineations may differ from the MOTUs based on DNA barcodes in the cases of young species or introgression. For example, the DNA barcodes of specimens identified as *Oreochromis niloticus* and *Sarotherodon galilaeus* showed small divergences and were grouped in the same MOTU by all analyses (ABGD, RESL, GMYC and bPTP, Fig. S1). These small distances also appear in a similar DNA barcoding study (Nwani et al., 2011). Representatives of these two closely-related genera, belonging to the same cichlid tribe (Dunz & Schliewen, 2013) are known to hybridise (Bezault et al., 2012). Hence, these observations could be the consequence of introgression even if we cannot exclude species misidentification, incomplete lineage sorting, taxonomic over-splitting or recent radiation (Nwani et al., 2011).

4.2 Conspecifics in adjacent ichthyological provinces (data set A)

Africa is divided into ichthyogeographical provinces on the basis of fish fauna composition (Roberts, 1975; Thieme et al., 2005), which reflects endemism and dispersal. Out of the 194 morphospecies collected here (496 specimens from CO and 245 from LG), only one was sampled in both ichthyological provinces and did not show high DNA barcode divergences among ichthyological

provinces (*Marcusenius moorii*, 0.7–1.7%). The other morphospecies were either morphologically different from those found in the neighbouring province or were represented by distinct clusters of DNA barcodes, showing divergences of 2.9–13%. Considering both morphology and DNA barcodes, we can say that 99.5% of all 194 morphospecies sampled in this study represent divergent lineages, illustrating the distinction of the ichthyofaunae of the two provinces.

4.3 DNA barcode divergences within ichthyological provinces (data set A)

DNA barcoding studies already suspected cases of intraspecific geographic resolution of COI at the river system level in southeastern Nigeria and in the north-eastern part of the Congo Basin (Decru et al., 2016; Nwani et al., 2011). Compared to the Congo Basin, which includes large tributary basins connected to the main Congo River, the Lower Guinean province includes several small to medium-sized coastal rivers (Brooks et al., 2011). This could explain why most cases of COI divergence observed within morphospecies (4/4) or species groups (4/5 morphospecies) were found in LG (Fig. 2B). Deep and uniform sampling is necessary to confirm the resolution at this level.

4.4 DNA barcoding as a joint effort (data set A+B)

Constructing large-scale reference libraries of DNA barcodes often requires the assembly of independently generated data sets. Combining our data set with that of Lowenstein et al. (2011) including species mainly collected in the same part of the Congolese drainage system confirmed the impressive ichthyological diversity of this region (Snoeks et al., 2011). The majority of the species collected in one study were not collected in the other study. This was measured both in terms of morphospecies (34/263 morphospecies in common) and DNA barcode clusters (42/261 putative species in common). These two measurements show that the error associated with alternative species name assignment has a minor effect on this observation and that much work remains to be

done to get a comprehensive knowledge of the fish diversity in the Lower Congo. Unfortunately, too few DNA barcoding studies intentionally report and discuss inconsistencies observed among reference libraries obtained for the same fauna. This is regrettable because inconsistencies and the way they are treated is crucial for the reliability of DNA barcoding species identifications. Our results show that species assignment becomes more problematic in the library combining data sets A and B (more inconsistencies and larger overlap between distances among and within morphospecies). First, misidentifications and mislabelling become obvious when misidentified specimens are compared with correctly identified specimens of the same species. This would explain the aberrant clustering of voucher specimen t-073-7242 (identified as *Garra ornata* in data set B, Cyprinidae) with specimens of a different family (A, Mochokidae). Mislabelling is also probable for voucher t-062-6197 (data set B), which was collected in DRC (according to its GenBank accession number HM418112) and identified as *Sanagia velifera*, an endemic species of Cameroon. The same is possible for specimens of *Labeo* (Fig. 5) and of *Monostichodus lootensi* (HM418231, Fig. 6). For more closely related species, when misidentification is more difficult to detect, more investigation is needed to judge if formally described species 1) were not consistently recognized by independent teams or 2) correspond to one single species or 3) cannot be distinguished on the basis of DNA barcodes. This could be the case for the following pairs of species showing identical or very similar DNA barcodes: *Labeo lineatus* (B) and *L. greenii* (A), *Mormyrops furcidens* (A) and *M. lineolatus* (B), *M. sirenoides* (A) and *M. masuianus* (B), *Pollimyrus nigripinnis* (A) and one haplotype of *P. maculipinnis* (B), *Citharinus macrolepis* (A) and *C. gibbosus* (B), *Enteromius holotaenia* (B) and *E. miolepis* (A). Anyway, most inconsistences between data sets were caused by a different way of treating unknown diversity. Indeed, in several cases, different lineages showing divergent DNA barcodes were identified with the only existing name available. This was likely the case for the following species, which were all described from other drainage systems or geographically distant places of the same rivers: *Clarias gabonensis* (Ogowe River, LG, Gabon), *Enteromius rubrostigma* (Ogowe River, LG, Gabon), *Opsaridium ubangiense* (Ubangi River, CO, Central African Republic) and

Eugnathichthys macroterolepis (Chiloango, LG, Angola). In other cases, morphological variants identified by us with provisional names (*Garra* cf. *ornata*, *Ctenopoma* cf. *nigropannosum*, *P.* cf. *nigripinnis*) do not show any DNA barcode divergence with the described species and correspond to either new phenotypes hitherto unknown for the species or recently diverged species.

Similarly to the study of Decru et al. (2016), the families showing most inconsistencies (Cyprinidae and Mormyridae) contain genera for which much of the species diversity remains overlooked. Taxonomic assignments are inevitably less accurate in the presence of undescribed lineages because species boundaries are unknown. DNA barcoding offers the opportunity to quantify the consistency of species delineations with a measure of genetic variability (Kress & Erickson, 2012). Our results highlight specific cases where the study of additional specimens and comparisons with type material are necessary to provide reliable species descriptions and it is evident that the developed DNA barcode reference library is far from comprehensive. It will become increasingly useful for biodiversity surveys when more species and more populations will be included. DNA barcoding ideally works as a species identification tool based on an established taxonomy (Desalle, 2006). This implies that the sequencing of new samples from unexplored areas is mainly useful to flag specimens with divergent DNA barcodes. Tools like ABGD, RESL, GMYC or bPTP provided very coherent sets of MOTUs in this study. We therefore recommend using provisional species assignments (i.e. not formal descriptions), for which MOTUs can be used in combination with morphospecies assignments to guide further taxonomic investigation when exploring poorly inventoried fauna.

Unfortunately, the rate-limiting factors in describing biodiversity will remain the collection of new specimens in the field and taxonomic revisions. This is especially true for Africa, where routine and geographically representative collection of species in the field remains problematic due to conflicts, inaccessibility and a lack of capacity and logistical support. Of particular concern are countries such as the DRC with high freshwater fish diversity and levels of endemism (Froese & Pauly, 2018; Swartz, Mwale, Hanner, & Swartz, 2008). The growing database of DNA barcodes indeed increases the speed and the consistency of fish identification by facilitating comparisons among reference collections

(Swartz et al., 2008). However, the rising number of putative new species reported on the basis of DNA barcode data will only speed up species descriptions if a considerable investment would be made in taxonomic experts that are capable to analyse them.

Acknowledgements

Fishes were collected and data obtained thanks to the Belgian Science Policy (BELSPO, Actie 1) project 'Multidisciplinair onderzoek op de diversiteit van de vissen van de Beneden-Kongo en de Pool Malebo', the FishBase for Africa project and project RA04F11 'Appui au laboratoire de biologie des populations de l'Université Marien Ngouabi, Brazzaville pour l'étude de la biodiversité et la conservation des poissons d'eaux douces du Congo-Brazzaville', both financed through the Framework Agreement between the RMCA and the Belgian Development Cooperation. For their Ph.D. projects, Soleil Wamuini and Armel Ibala-Zamba benefitted from a sandwich grant from the Belgian Development Cooperation. DNA barcoding was funded by the Belgian Science Policy (BELSPO). The authors thank the three reviewers of this article for their very constructive suggestions.

References

African FISH-BOL Regional Working Group. (2008). *Workshop report of the first African Fish Barcode of Life initiative (FISH-BOL) Regional Working Group*. Addis Ababa, Ethiopia, 21 September 2008. Retrieved from http://www.barcodeoflife.org/sites/default/files/materials/FISH-BOLPAFFAreport.pdf

April, J., Mayden, R., Hanner, R., & Bernatchez, L. (2011). Genetic calibration of species diversity among North America 's freshwater fishes. *Proceedings of the National Academy of Sciences*, *108*, 10602–10607. doi: 10.1073/pnas.1016437108

Becker, S., Hanner, R., & Steinke, D. (2011). Five years of FISH-BOL: Brief status report. *Mitochondrial DNA*, *22*(sup1), 3–9. doi: 10.3109/19401736.2010.535528

Berrebi, P., Chenuil, A., Kotlík, P., Machordom, A., & Tsigenopoulos, C. S. (2014). Disentangling the evolutionary history of the genus *Barbus* sensu lato, a twenty years adventure. In M. J. Alves, A. Cartaxana, A. M. Correia, & L. F. Lopes (Eds.), *Professor Carlos Almaça (1934–2010) – Estado da arte em áreas científicas do seu interesse* (pp. 29–55). Lisboa, Portugal: Museu Nacional de História Natural e da Ciência.

Bezault, Z., Rognon, X., Clota, F., Gharbi, K., Baroiller, J-F., & Chevassus, B. (2012). Analysis of the meiotic segregation in intergeneric hybrids of Tilapias. *International Journal of Evolutionary Biology*, *2012*, 817562. doi: 10.1155/2012/817562

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., … Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology, 10(4), e1003537. doi: 10.1371/journal.pcbi.1003537

Boulenger, G. A. (1901). *Les poissons du bassin du Congo*. Brussels, Belgium: Publication de l'État Indépendant du Congo.

Brooks, E. G. E., Allen, D. J., & Darwall, W. R. T. (2011). *The Status and Distribution of Freshwater Biodiversity in Central Africa*. Cambridge, United Kingdom and Gland, Switzerland: International Union for Conservation of Nature and Natural Resources (IUCN).

Chakrabarty, P. (2010). Genetypes: a concept to help integrate molecular phylogenetics and taxonomy. *Zootaxa 2632*, 67–68.

Collins, R. A., Boykin, L. M., Cruickshank, R. H., & Armstrong, K. F. (2012). Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods in Ecology and Evolution*, *3*, 457–465. doi: 10.1111/j.2041-210X.2011.00176.x

Darwall, W. R. T., Smith, K. G., Allen, D. J., Holland, R. A., Harrison, I. J., & Brooks, E. G. E. (2011). *The Diversity of Life in African Freshwaters: Under Water, Under Threat. An analysis of the status and distribution of freshwater species throughout mainland Africa*. Cambridge, United Kingdom and Gland, Switzerland: International Union for Conservation of Nature and Natural Resources (IUCN).

Decru, E., Moelants, T., De Gelas, K., Vreven, E., Verheyen, E., & Snoeks, J. (2016). Taxonomic challenges in freshwater fishes: A mismatch between morphology and DNA barcoding in fish of the north-eastern part of the Congo basin. *Molecular Ecology Resources*, *16*(1), 342–352. doi: 10.1111/1755-0998.12445

Decru, E., Snoeks, J., & Vreven, E. (2015). Taxonomic evaluation of the *Hepsetus* from the Congo basin with the revalidation of *H. microlepis* (Teleostei: Hepsetidae). *Ichthyological Exploration of Freshwaters*, *26*(3), 273–287.

Decru, E., Vreven, E., Danadu, C., Walanga, A., Mambo, T., & Snoeks, J. (2017). Ichthyofauna of the Itimbiri, Aruwimi, and Lindi/Tshopo rivers (Congo basin): Diversity and distribution patterns. *Acta Ichthyologica et Piscatoria*, *47*(3), 225–247. doi: 10.3750/AIEP/02085

Decru, E., Vreven, E., & Snoeks, J. (2012). A revision of the West African *Hepsetus* (Characiformes: Hepsetidae) with a description of *Hepsetus akawo* sp. nov. and a redescription of *Hepsetus*

odoe (Bloch, 1794). *Journal of Natural History*, *46*(1–2), 1–23. doi: 10.1080/00222933.2011.622055

Decru, E., Vreven, E., & Snoeks, J. (2013). A revision of the Lower Guinean *Hepsetus* species (Characiformes; Hepsetidae) with the description of *Hepsetus kingsleyae* sp. nov. *Journal of Fish Biology*, *82*(4), 1351–1375. doi: 10.1111/jfb.12079

Desalle, R. (2006). Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. *Conservation Biology*, *20*(5), 1545–1547. doi: 10.1111/j.1523-1739.2006.00543.x

Dunz, A. R. & Schliewen, U. K. (2013). Molecular phylogeny and revised classification of the haplotilapiine cichlid fishes formerly referred to as "*Tilapia*". *Molecular Phylogenetics and Evolution*, *68*(1), 64–80. doi: 10.1016/j.ympev.2013.03.015

Dunz, A. R., Vreven, E., & Schliewen, U. K. (2012). *Congolapia*, a new cichlid genus from the central Congo basin (Perciformes: Cichlidae). *Ichthyological Exploration of Freshwaters*, *23*(2), 155–179.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299.

Froese, R., & Pauly, D. (2018, January 25). FishBase. World Wide Web electronic publication. Retrieved February 12, 2017, from www.fishbase.org

Hanner, R., Desalle, R., Ward, R. D., & Kolokotronis, S.-O. (2011). The Fish Barcode of Life (FISH-BOL) special issue. *Mitochondrial DNA*, *22*(sup1), 1–2. doi: 10.3109/19401736.2011.598767

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *270*(1512), 313–321. doi: 10.1098/rspb.2002.2218

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

Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., Burridge, M., … Bernatchez, L. (2008). Identifying Canadian freshwater fishes through DNA Barcodes. *PLoS ONE*, *3*(6), e2490. doi: 10.1371/journal.pone.0002490

Janzen, D. H., Hallwachs, W., Blandin, P., Burns, J. M., Cadiou, J. M., Chacon, I., … Wilson, J. J. (2009). Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources*, *9*(sup1), 1–26. doi: 10.1111/j.1755-0998.2009.02628.x

Kress, W. J., & Erickson, D. L. (2012). DNA Barcodes: Methods and Protocols. In *Methods in Molecular Biology* (pp. 3–8). Totowa, New Jersey: Humana Press. doi: 10.1007/978-1-61779-591-6_1

Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, *14*, 82. doi: 10.1186/1471-

2148-14-82Lévêque, C., Oberdorff, T., Paugy, D., Stiassny, M. L. J., & Tedesco, P. A. (2008). Global diversity of fish (Pisces) in freshwater. *Hydrobiologia*, *595*, 545–567. doi: 10.1007/978-1-4020-8259-7

Lowenstein, J. H., Osmundson, T. W., Becker, S., Hanner, R., & Stiassny, M. L. J. (2011). Incorporating DNA barcodes into a multi-year inventory of the fishes of the hyperdiverse Lower Congo River, with a multi-gene performance assessment of the genus Labeo as a case study. *Mitochondrial DNA*, *22*(sup1), 52–70. doi: 10.3109/19401736.2010.537748

Meier, R., Shiyang, K., Vaidya, G., & Ng, P. K. L. (2006). DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, *55*(5), 715–728.

Messing, J., Crea, R., & Seeburg, P. H. (1981). A system for shotgun DNA sequencing. *Nucleic Acids Research*, *9*(2), 309–321. doi:

Miller, M. A., Pfeiffer, W. & Schwartz, T. (2010, November 14). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Paper presented at the 2010 Gateway Computing Environments Workshop (GCE), pp. 1–8. doi: 10.1109/GCE.2010.5676129

Nascimento, M. H. S., Almeida, M. S., Veira, M. N. S., Limeira Filho, D., Lima, R. C., Barros, M. C., & Fraga, E. C. (2016). DNA barcoding reveals high levels of genetic diversity in the fishes of the Itapecuru Basin in Maranhão, Brazil. *Genetics and Molecular Research*, *15*(3). doi: 10.4238/gmr.15038476

Nwani, C. D., Becker, S., Braid, H. E., Ude, E. F., Okogwu, O. I., & Hanner, R. (2011). DNA barcoding discriminates freshwater fishes from southeastern Nigeria and provides river system-level phylogeographic resolution within some species. *Mitochondrial DNA*, *22*(sup1), 43–51. doi: 10.3109/19401736.2010.536537

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, *20*(2), 289–290. doi: 10.1093/bioinformatics/btg412

Pereira, L. H. G., Pazian, M. F., Hanner, R., Foresti, F., & Oliveira, C. (2011). DNA barcoding reveals hidden diversity in the Neotropical freshwater fish *Piabina argentea* (Characiformes: Characidae) from the Upper Paran Basin of Brazil. *Mitochondrial DNA*, *22*(sup1), 87–96. doi: 10.3109/19401736.2011.588213

Pereira, L. H., Hanner, R., Foresti, F., & Oliveira, C. (2013). Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC Genetics*, *14*(1), 20. doi: 10.1186/1471-2156-14-20

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., & Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology, 55*, 595–609. doi: 10.1080/10635150600852011

Pugedo, M. L., de Andrade Neto, F. R., Pessali, T. C., Birindelli, J. L. O., & Carvalho, D. C. (2016). Integrative taxonomy supports new candidate fish species in a poorly studied neotropical region: the Jequitinhonha River Basin. *Genetica*, *144*(3), 341–349. doi: 10.1007/s10709-016-

9903-4

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, *21*(8), 1864–77. doi: 10.1111/j.1365-294X.2011.05239.x

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD : The Barcode of Life Data System. *Molecular Ecology Notes*, *7*, 355–364. doi: 10.1111/j.1471-8286.2006.01678.x

Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-based registry for all animal species: the Barcode Index Number (BIN) System. *PLoS ONE*, *8*(7), e66213. doi: 10.1371/journal.pone.0066213

Roberts, T. R. (1975). Geographical distribution of African freshwater fishes. *Zoological Journal of the Linnean Society*, *57*(4), 249–319. doi: 10.1111/j.1096-3642.1975.tb01893.x

Roberts, T. R., & Stewart, D. J. (1976). an Ecological and Systematic Survey of Fishes in the Rapids of the Lower Zaire or Congo River. *Bulletin of the Museum of Comparative Zoology*, *147*(6), 239–317.

Shen, Y., Guan, L., Wang, D., & Gan, X. (2016). DNA barcoding and evaluation of genetic diversity in Cyprinidae fish in the midstream of the Yangtze River. *Ecology and Evolution*, *6*(9), 2702–2713. doi: 10.1002/ece3.2060

Sheth, B. P., & Thaker, V. S. (2017). DNA barcoding and traditional taxonomy: an integrated approach for biodiversity conservation. *Genome*, *60*(7), 618–628. doi: 10.1139/gen-2015-0167

Shumway, C., Mosibono, D., Ifuta, S., Sullivan, J., Schelly, R., Punga, J., … Puema, V. (2003). *Biodiversity Survey: Systematics, Ecology and Conservation Along the Congo River, Congo River Environment and Development Project (CREDP)*. Boston, Massachusetts: New England Aquarium

Snoeks, J., Harrison, I. J., & Stiassny, M. L. J. (2011). The status and distribution of freshwater fishes. In *The diversity of life in African freshwaters: Under Water, Under Threat. An analysis of the status and distribution of freshwater species throughout mainland Africa* (pp. 42–73). Cambridge, United Kingdom and Gland, Switzerland: International Union for Conservation of Nature and Natural Resources (IUCN).

Srivathsan, A., & Meier, R. (2012). On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature, *Cladistics*, *28*, 190–194. doi: 10.1111/j.1096-0031.2011.00370.x

Stamatakis, A. (2015). Using RAxML to Infer Phylogenies. *Current protocols in bioinformatics, 51*, 6.14.1–6.14.14. doi: 10.1002/0471250953.bi0614s51

Stiassny, M. L. J., & Mamonekene, V. (2007). *Micralestes* (Characiformes, Alestidae) of the lower Congo River, with a in the Democratic Republic of Congo. *Zootaxa*, *29*, 17–29. doi: 10.5281/zenodo.179053

Stiassny, M. L. J., Teugels, G. G., & Hopkins, C. D. (2007). *Poissons d'eaux douces et saumâtres de*

basse Guinée, ouest de l'Afrique centrale, volumes 1 and 2. Paris, France: IRD Éditions.

Swartz, E. R., Mwale, M., Hanner, R., & Swartz, E. R. (2008). A role for barcoding in the study of African fish diversity and conservation . S Afr J Sci A role for barcoding in the study of African fish diversity and conservation. *South African Journal of Science*, *104*, 293–298.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729. doi: 10.1093/molbev/mst197

Thieme, M. L., Abell, R., Burgess, N., Lehner, B., Dinerstein, E., Olson, D., … Skelton, P. (2005). *Freshwater ecoregions of Africa and Madagascar: A conservation assessment*. Washington, USA: Island Press.

Tsigenopoulos, C. S., Kasapidis, P., & Berrebi, P. (2010). Phylogenetic relationships of hexaploid large-sized barbs (genus *Labeobarbus*, Cyprinidae) based on mtDNA data. *Molecular Phylogenetics and Evolution*, *56*(2), 851–856. doi: 10.1016/j.ympev.2010.02.006

Van Ginneken, M., Decru, E., Verheyen, E., & Snoeks, J. (2017). Morphometry and DNA barcoding reveal cryptic diversity in the genus *Enteromius* (Cypriniformes: Cyprinidae) from the Congo basin, Africa – Corrigendum. *European Journal of Taxonomy*, (314). doi: 10.5852/ejt.2017.314

Van Steenberge, M., Gajdzik, L., Chilala, A., Snoeks, J., & Vreven, E. (2017). Don't judge a fish by its fins: species delineation of Congolese *Labeo* (Cyprinidae). *Zoologica Scripta*, *46*(3), 264–274. doi: 10.1111/zsc.12203

Van Steenberge, M., Vreven, E., & Snoeks, J. (2014). The fishes of the Upper Luapula area (Congo basin): A fauna of mixed origin. *Ichthyological Exploration of Freshwaters*, *24*(4), 329–345.

Vreven, E. J. W. M. N., Musschoot, T., Snoeks, J., & Schliewen, U. K. (2016). The African hexaploid Torini (Cypriniformes: Cyprinidae): review of a tumultuous history. *Zoological Journal of the Linnean Society*, *177*(2), 231–305. doi: 10.1111/zoj.12366

Walsh, G., & Mamonekene, V. (2014). A collection of fishes from tributaries of the lower Kouilou, Noumbi and smaller coastal basin systems, Republic of the Congo, Lower Guinea, west-central Africa. *Check List*, *10*(4), 900–912. doi: 10.15560/10.4.900

Wamuini, S. L., Vreven, E., Vandewalle, P., Mutambue, S., & Snoeks, J. (2010). Contribution to the knowledge of the ichthyofauna in the Inkisi River, Lower Congo (RDC). *Cybium: international journal of ichthyology*, *34*, 83–91.

Ward, R. D., Hanner, R., & Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, *74*(2), 329–356. doi: 10.1111/j.1095-8649.2008.02080.x

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. N. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1847–1857. doi: 10.1098/rstb.2005.1716

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with

applications to phylogenetic placements. *Bioinformatics*, *29*(22), 2869–2876. doi: 10.1093/bioinformatics/btt499

Data Accessibility Statement

DNA sequences: Process ID in the Barcode of Life Data Systems (BOLD) are from BCOVR001-17 to BCOVR741-17 (project BCOVR) and accession numbers in GenBank are from MK073961 to MK074701. Hologenetype of *Synodontis carineae*: MK074654. Paragenetypes of *Synodontis carineae*: MK074655-8 and MK074660.

DNA alignments of the COI data sets A and A+B are deposited in the Dryad Digital Repository: https://doi.org/10.5061/dryad.5qj120q

Author Contributions

EVer, JS, EVre, JVH, ZTN, FCB and GS designed the research. MVS, EVre, GB, MH, SWL, AIZ, VM, TM and ED collected the fishes and performed the morphological identifications. FCB, GS, ZTN and JVH performed the lab work. GS, EVer, ZTN and JVH analysed the data. GS, EVer, ZTN, JS, KJ, EVre and MVS wrote the manuscript.

Figures and tables

Figure 1. Sampling locations of this study and Lowenstein et al. (2011) at different scales (A, B, C). Orange symbols mark the collection sites of this study situated within the Congolese ichthyological province (diamonds: Luki, stars: Inkisi, disks: Djoué and Pool Malebo, triangles: Lefini). Red symbols represent our sampling records from the Lower Guinean ichthyological province (disks: Kouilou-Niari, diamonds: Nyanga, star: Polo Ogowe and triangle: Ngongo). Black dots represent sampling records of Lowenstein et al. (2011). Blue lines and orange lines represent rivers and political borders, respectively.

Figure 2. A: Boxplots representing the ranges of distances observed in data sets A (left) and A+B within and among morphospecies (top), and within and among MOTUs determined using ABGD (bottom) (right). This was done for all families and for the five families represented by more than 60 sequences (Alestidae, Cichlidae, Cyprinidae, Distichodontidae and Mormyridae). B: Ranges of distances among specimens of the same morphospecies collected within and among river basins. CO: Congolese ichthyological province, LG: Lower Guinean ichthyological province. Coloured boxes represent the interquartile range. Bold horizontal bars represent medians. Whiskers represent the range of values situated <1.5 times the interquartile range from the box. Open circles represent outliers outside this range.

Figure 3**.** A: Subtrees showing distances within morphospecies found in both ichthyological provinces. B: Subtrees showing distances between pairs of similar morphospecies collected in neighbouring provinces. These subtrees were extracted from the neighbour-joining tree based on the new barcode data set (data set A) and using uncorrected p-distances. Bootstrap values are given at the nodes.

Figure 4**.** A: Subtrees showing distances among relatively similar morphospecies collected in the same ichthyological province. B: Subtrees showing distances within morphospecies collected in different rivers of the same ichthyological province. C: Subtrees showing distances within morphospecies collected in the same river. These subtrees were extracted from the neighbour-joining tree based on the new barcode data set (data set A) and using uncorrected p-distances. Bootstrap values are given at the nodes.

Figure 5. Neighbour-Joining (NJ) tree (uncorrected p-distance) for all DNA barcode sequences of the family Alestidae. Bootstrap values are given at the nodes. (A): new records from this study; (B): records from Lowenstein et al. (2011).

Figure 6. Neighbour-Joining (NJ) tree (uncorrected p-distance) for all DNA barcode sequences of the family Cichlidae. Bootstrap values are given at the nodes. (A): new records from this study; (B): records from Lowenstein et al. (2011). Bold annotations indicate incongruences between DNA barcode clusters in the tree and morphospecies assignments (S: morphospecies split in different clusters; L: different morphospecies lumped in the same cluster).

Figure 7. Neighbour-Joining (NJ) tree (uncorrected p-distance) for all DNA barcode sequences of the family Cyprinidae. Bootstrap values are given at the nodes. (A): new records from this study; (B): records from Lowenstein et al. (2011). Bold annotations indicate incongruences between DNA barcode clusters in the tree and morphospecies assignments (S: morphospecies split in different clusters; L: different morphospecies lumped in the same cluster; *: incongruence obtained after merging data sets A and B).

Figure 8. Neighbour-Joining (NJ) tree (uncorrected p-distance) for all DNA barcode sequences of the family Distichodontidae. Bootstrap values are given at the nodes. (A): new records from this study; (B): records from Lowenstein et al. (2011). Bold annotations indicate incongruences between DNA barcode clusters in the tree and morphospecies assignments (S: morphospecies split in different clusters; L: different morphospecies lumped in the same cluster; *: incongruence obtained after merging data sets A and B).

Figure 9. Neighbour-Joining tree (uncorrected p-distance) for all DNA barcode sequences of the family Mormyridae. Bootstrap values are given at the nodes. (A): new records from this study; (B): records from Lowenstein et al. (2011). Bold annotations indicate incongruences between DNA barcode clusters in the tree and morphospecies assignments (S: morphospecies split in different clusters; L: different morphospecies lumped in the same cluster; *: incongruence obtained after merging data sets A and B).

Supporting information

Table S1 (sup-001-TableS1-S2). List of all fishes analysed in this study (data sets A, B and A+B), with information on their morphological identification (according to the taxonomy of 2012 and 2018), their collection and their molecular operational taxonomic unit ID based on different methods (ABGD, RESL, GMYC and bPTP).

Table S2 (sup-001-TableS1-S2). List of all fish species analysed in this study, with number of specimens sampled per species in the different data sets (A, B and A+B) and information on the consistency between (morpho)species identifications (according to the taxonomy of 2012 and 2018) and clusters in the neighbour-joining (NJ) tree. OK: consistent, -: no data, split: morphospecies split in different clusters of the NJ tree, lumped: different morphospecies lumped in one cluster of the NJ tree, new: new inconsistency observed after merging data sets A and B.

Figure S1 (sup-002-FigS1-S2). Neighbour joining tree constructed with pairwise deletion and using pairwise uncorrected p-distances among the DNA barcode sequences (5'-end of the cytochrome c oxidase subunit I gene, 589-652 bp) of the 741 fishes sequenced here (data set A). Bootstrap values are indicated at nodes (1000 replicates).

Figure S2 (sup-002-FigS1-S2). Neighbour joining tree constructed with pairwise deletion and using pairwise uncorrected p-distances among the DNA barcode sequences (5'-end of the cytochrome c oxidase subunit I gene, 589-652 bp) of 932 fishes sequenced here (data set A, 741 specimens) and by Lowenstein et al. (2011) (data set B, 191 specimens). Bootstrap values are indicated at nodes (1000 replicates).