Received: 6 September 2019 | Revised: 5 March 2020 | Accepted: 11 March 2020

DOI: 10.1111/jbi.13858

RESEARCH PAPER

Journal of Biogeography | WILEY

Unravelling the evolution of Africa's drainage basins through a widespread freshwater fish, the African sharptooth catfish *Clarias gariepinus*

Maarten W. Van Steenberge[1,2,3,4,5] | Maarten P. M. Vanhove[1,5,6,7] | Auguste Chocha Manda[1,8,9] | Maarten H. D. Larmuseau[1,10] | Belinda L. Swart[11] | Faustin Khang'Mate[8] | Allan Arndt[12] | Bart Hellemans[1] | Jeroen Van Houdt[13] | Jean-Claude Micha[9] | Stephan Koblmüller[4] | Rouvay Roodt-Wilding[11] | Filip A. M. Volckaert[1]

[1]Laboratory of Biodiversity and Evolutionary Genomics, Laboratory of Socioecology and Social Evolution & Department of Human Genetics, KU Leuven, Leuven, Belgium

[2]Ichthyology Group, Royal Museum for Central Africa, Tervuren, Belgium

[3]Operational Directorate Taxonomy and Phylogeny, Royal Belgian Institute for Natural Sciences, Brussels, Belgium

[4]Institute of Biology, University of Graz, Graz, Austria

[5]Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic

[6]Research Group Zoology: Biodiversity and Toxicology, Centre for Environmental Sciences, Hasselt University, Diepenbeek, Belgium

[7]Zoology Unit, Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland

[8]Research Unit of biodiversity and sustainable wetland exploitation, Faculty of Agronomy & Research Unit of reproduction, Faculty of Veterinary Sciences, University of Lubumbashi, Lubumbashi, Democratic Republic of the Congo

[9]Research Unit in Environmental and Evolutionary Biology, University of Namur, Namur, Belgium

[10]Histories VZW, Mechelen, Belgium

[11]Department of Genetics, Stellenbosch University, Stellenbosch, Republic of South Africa

Abstract

Aim: The formation history of Africa's current river basins remains largely unknown. In order to date changes in landscape and climate, we studied the biogeography of the African freshwater fish with the largest natural distribution. We also validated biogeographical units.

Location: Continental Africa.

Taxon: *Clarias gariepinus* sl.

Methods: We investigated mitochondrial *cytb* sequences of 443 individuals from 97 localities, using a haplotype network and a genetic landscape analysis. We inferred a dated phylogeny using maximum likelihood and Bayesian inference approaches and reconstructed ancestral areas with S-DEC and S-DIVA models. Microsatellite genotyping complemented the mitochondrial approach in the Congo basin, where the latter revealed complex patterns.

Results: Limited differentiation is found in northern and south-western Africa, and sharp genetic differentiation in the continent's east and centre. Populations with affinities to neighbouring basins occur at the edges of the Congo province. High diversity exists in the south of the Congo basin. The Zambezi province is partitioned into eastern, central and western sectors. In the east, specimens were related to those from the Congo. In the west, they were similar to Southern representatives. Phylogenetic inference placed the origin of *C. gariepinus* in the East Coast, with intraspecific diversification starting around the Great Lakes. These events occurred ca. 4.8–1.65 and 2.3–0.8 MYA respectively.

Main conclusions: Clades of *C. gariepinus* sl. show a clear geographical signature. The origin of *C. gariepinus* in the East Coast and diversification around the Great Lakes

[12]Department of Biology, University of the Fraser Valley, Abbotsford, Canada

[13]Research and Development, Janssen Pharmaceutica NV, Beerse, Belgium

Correspondence
Maarten Van Steenberge, Royal Belgian Institute for Natural Sciences Vautierstraat 29, 1000 Brussels, Belgium.
Email: mvansteenberge@naturalsciences.be

Funding information
Fonds Wetenschappelijk Onderzoek; Belgisch Ontwikkelingsagentschap

Handling Editor: Sergio Floeter

coincided with the periods of increased aridity. Low genetic differentiation in northern and southern Africa may result from connectivity during recent periods of higher rainfall. In contrast to other widespread African freshwater fish, colonization rather than extinction seemed to mediate distribution patterns. This can be explained by a high ecological tolerance. We highlight the species' suitability to study landscape and climate evolution at various scales.

KEYWORDS
Africa, Clariidae, fish, ichthyofaunal provinces, phylogeography, river basin

1 | INTRODUCTION

Africa has a complex hydrology that is characterized by numerous inland deltas, palaeolakes, cataracts and elbows of river captures. All are reminders of a highly dynamic past (Goudie, 2005). Even for the largest rivers on the continent, the question remains how and when they obtained their present-day form. Major large-scale geological events such as the formation of the central African shear zone (Fairhead, 1988), the East African rift (Stankiewicz & de Wit, 2006) and the Kalahari uplift (Cotteril & de Wit, 2011) were pivotal in shaping the boundaries of the current drainage patterns. Tracing the origin and the timing of associated changes in hydrology remains difficult when solely relying on the stratigraphic record (Watchman & Tweddle, 2002). Yet, the evolutionary histories of aquatic organisms can provide proxies to identify, quantify and date changes in hydrology (Burridge, Craw, Jack, King, & Waters, 2008; Skelton, 1994).

Cycling between dry and wet periods in the Pleistocene and Pliocene impacted the evolution of Africa's landscapes and fauna, including early hominids (deMenocal, 2014; Maslin et al., 2014). This climate cycling led to alternate expansions and contractions of savanna- and forest-like habitats, in parallel with alterations of deep- and low-water stands in Africa's Great Lakes (Malinsky & Salzburger, 2016). In tropical rivers, such changes in climate triggered the instances of immigration, extinction and allopatric divergence (Tedesco, Oberdorff, Lasso, Zapata, & Hugueny, 2005), creating the current fish faunas. Using similarities between fish faunas allowed ichthyologists to divide Africa into ten ichthyofaunal provinces (Snoeks & Getahun, 2013). These are: the Congo, East Coast, Lower Guinea, Maghreb, Malagasy, Nilo-Sudan, Quanza, Southern, Upper Guinea and Zambezi provinces (Figure 1). When taking ecological similarity into account, a more fine-scale subdivision of the continent in 95 freshwater ecoregions has also been proposed (Thieme et al., 2005).

Aquatic organisms have been used to unravel the evolutionary history of watersheds both at regional (Burridge et al., 2008) and continental (Goodier, Cotterill, O'Ryan, Skelton, & de Wit, 2011; Pinton, Agnèse, Paugy, & Otero, 2013) scales. As the effects of changes in hydrology depend on the ecology of a species, studies of organisms that occupy different niches will highlight other aspects of drainage evolution. Previous studies on pan-African or southern African scales focussed on aquatic Nile crocodile, *Crocodylus niloticus* Laurenti, 1768, (Cotterill & Goodier, 2008), swamp-dwelling lechwe antelopes *Kobus* Smith, 1840 (Cotterill & Goodier, 2008), viviparid freshwater snails (Schultheiss, Van Boxlaer, Riedel, von Rintelen, & Albrecht, 2014) and, for fish, on squeaker catfish *Synodontis* Cuvier, 1816 (Day et al., 2013), spiny eels *Mastacembelus* Scopoli, 1777 (Day et al., 2017) and tigerfish *Hydrocynus* Cuvier, 1816 (Alestidae; Goodier et al., 2011).

Clarias gariepinus (Burchell, 1822) is the African freshwater fish with the largest (natural) distribution (Skelton, 2001). It is a true generalist that occurs in almost all freshwater systems throughout continental Africa, except for the extreme north- and south-west, and in the Levant and southern Anatolia, and is thus well suited to validate the boundaries of biogeographical entities using genetic data (Arndt et al., 2003; Giddelo, Arndt, & Volckaert, 2002; Rognon et al., 1998). *Clarias gariepinus* has an omnivorous diet, is tolerant of a wide range of water temperatures and is often ecologically dominant. It endures harsh conditions and tolerates high turbidity, low levels of oxygen and desiccation. The species has remarkable drought adaptations such as a secondary suprabranchial organ that allows it to take up atmospheric oxygen. It can move overland to escape the driest conditions and is frequently the last or only species inhabiting diminishing pools, poorly oxygenated swamps or drying rivers (Skelton, 2001). The species is rendered paraphyletic by both *C. anguillaris* (Linnaeus, 1759), which has a similar ecology as *C. gariepinus*, and by the nine species of *Bathyclarias* Jackson, 1959 (Angnèse & Teugels, 2001; Rognon et al., 1998), which became adapted to deep water conditions. All of these are fully sympatric with *C. gariepinus*, with the former being widespread in western and northern Africa, and the latter nine being endemic to Lake Malawi. Although it cannot be excluded that *C. gariepinus* contains undescribed variation, for the phylogeographical study presented here, working with a monophyletic lineage is more important than with a taxonomic unit. Hence, we will refer to the assemblage of *C. gariepinus*, *C. anguillaris* and the species flock of *Bathyclarias* as *C. gariepinus sensu lato* (sl.).

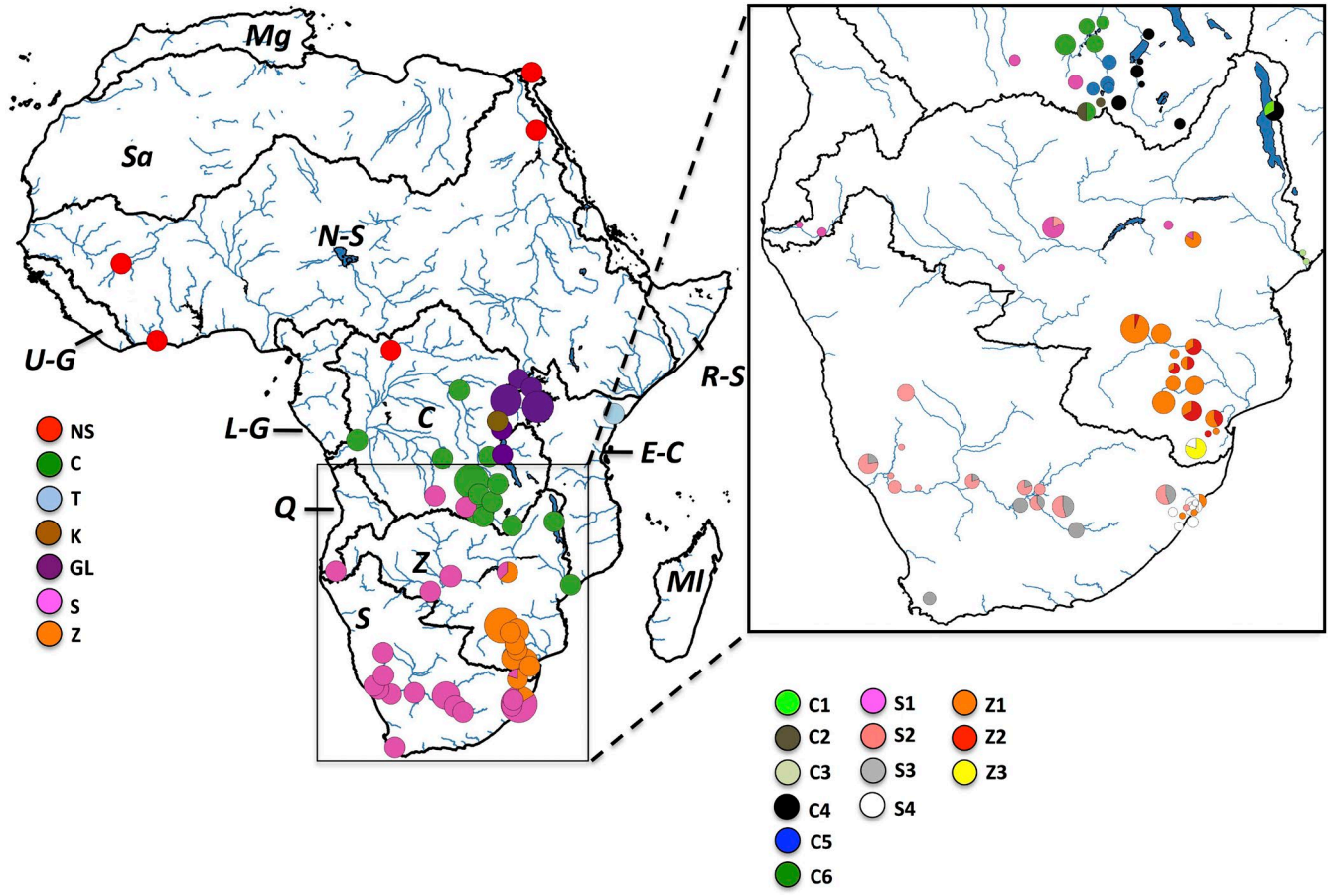

FIGURE 1 Map of sampling locations of the specimens of *C. gariepinus* sl. used in the study (left), including an out-take of southern Africa (right). Colours denote the clusters identified in the haplotype network to which the specimens belong. On the detailed map of southern Africa, colours denote the subclusters that were identified in clusters C, S and Z (see Figure 2). Sizes of vertices denote the number of samples from a certain location. On the left-hand figure, locations from a single continuous water body, within a distance of less than 100 km were grouped. Boundaries denote the ichthyofaunal provinces following Snoeks and Getahun (2013), with: C: Congo, E-C: East Coast, L-G: Lower Guinea, Mg: Maghreb, MI: Malagasy, N-S: Nilo-Sudan, Q: Quanza, S: Southern, U-G: Upper Guinea, Z: Zambezi. The arid regions near the Red Sea (R-S) and in the Sahara (S) do not belong to any ichthyofaunal province. *Clarias gariepinus* sl. is native to all of continental Africa's ichthyofaunal provinces, except for the Maghreb [Colour figure can be viewed at wileyonlinelibrary.com]

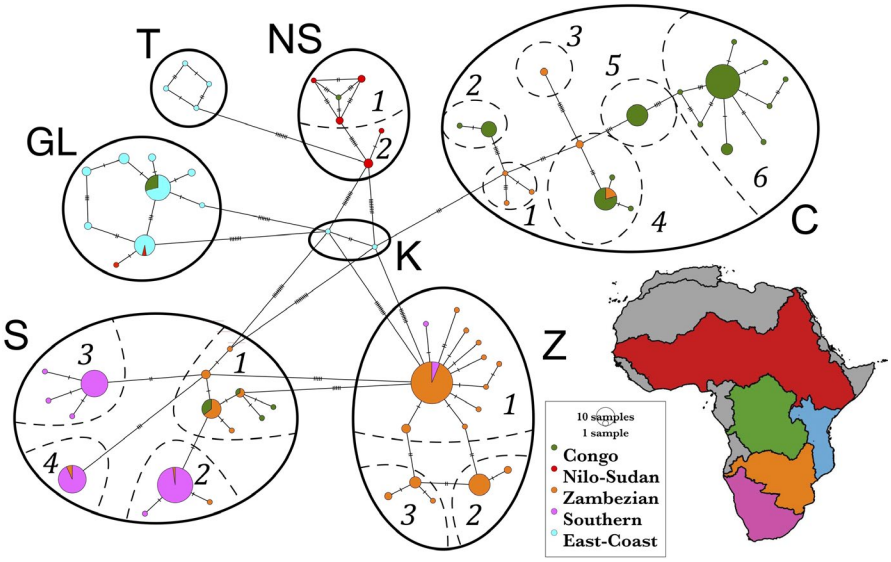

FIGURE 2 TCS haplotype network of the *cyt*b sequences of *C. gariepinus* sl. Seven clusters were identified in the network (letter codes), some of which were further divided into subclusters (numbers). Vertices were coloured based on the ichthyofaunal provinces from which the samples originate as indicated on the map to the right, their sizes denote the number of samples. Using the dominant origin of their haplotypes, clusters were labelled as: Nilo-Sudan (NS), Congo (C), Kivu (K), Great Lakes (GL), Southern (S), Tana (T) and Zambezi (Z) [Colour figure can be viewed at wileyonlinelibrary.com]

While the evolutionary history of *Hydrocynus*, which are restricted to warm, well-oxygenated, large rivers and lakes, is useful to track changes in the topology of major river channels (Goodier et al., 2011), the history of *C. gariepinus* will also tell the story of swamplands, headwaters and relict populations. Additionally, while the isolation led to speciation within *Hydrocynus*, *Synodontis* and *Mastacembelus*, this was prevented in *C. gariepinus* by its niche width and dispersal abilities. This lack of speciation allowed signatures of subsequent biogeographical events to become embedded in the genomes of single populations.

We infer a continental mitochondrial phylogeography of *C. gariepinus* sl., and provide a regional microsatellite distribution pattern to test the biogeographical potential of *C. gariepinus* with two evolutionary hypotheses: (1) the current genetic structure of the species can be used to validate biogeographical entities and (2) its phylogenetic history can be used to date and reconstruct patterns of landscape evolution and climatic events.

2 | MATERIALS AND METHODS

2.1 | Data collection

We sequenced fish sampled at 97 sites across Africa (Table S1; Figure 1), including those analysed for other markers by Giddelo et al. (2002) and Arndt et al. (2003). Fin clips were preserved at collection sites in absolute ethanol or in salt-saturated dimethylsulphoxide. We extracted DNA using proteinase *K* and CTAB buffer or using the NucleoSpin Tissue kit (Macherey-Nagel) and amplified the mitochondrial cytochrome *b* (*cyt*b) gene using the L14724/H15915 primer combination (Irwin, Kocher, & Wilson, 1991) and the following protocol: 94°C for 60 s, followed by 30 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 40 s and final extension at 72°C for 5 s. We added all sequences on NCBI GenBank for which data on field-based collection were available (AF126823-24; c [seven sequences, 11 specimens] AF235933; AF475153; DQ646360-72) to the dataset of newly generated sequences (MN941435-185) and obtained 443 sequences of *C. gariepinus* (*N* = 407), *C. anguillaris* (*N* = 3), the species flock of *Bathyclarias* (*N* = 8) and *C. ngamensis* (*N* = 25). Sequences originate from the Congo (*N* = 116 *C. gariepinus*, 18 *C. ngamensis*), East Coast (*N* = 49 *C. gariepinus*), Nilo-Sudan (*N* = 11 *C. gariepinus*, 3 *C. anguillaris*), Southern (*N* = 109 *C. gariepinus*) and Zambezi (*N* = 122 *C. gariepinus*, 8 *Bathyclarias* spp., 7 *C. ngamensis*) ichthyofaunal provinces. Maps were made using QGIS (QGIS Development Team, 2009)

2.2 | Haplotype network

We applied the TCS approach (Templeton, Crandall, & Sign, 1992) in PopART (Leigh & Bryant, 2015) to generate a haplotype network, using all 418 sequences belonging to *C. gariepinus* sl.: *C. gariepinus* (*N* = 407), *C. anguillaris* (*N* = 3) and *Bathyclarias* spp. (*N* = 8). We grouped specimens according to origin, that is, ichthyofaunal provinces (Snoeks & Getahun, 2013; Table S1). We identified clusters and subclusters in the network, based on a combination of geographical and genetic distinctness (Table S2).

2.3 | Genetic landscape shape interpolation

We investigated the spatial patterns of genetic differentiation by genetic landscape shape interpolation analysis implemented in Alleles in Space v1.0 (Miller, 2005). This analysis is based on a Delaunay triangulation connectivity network in which residual genetic distances are derived from a linear regression of genetic versus geographic distances (recommended for datasets with substantial variation in distances between sites; Manni, Guerard, & Heyer, 2004). We set grid size to 1 × 1 degrees, and used a distance weighting parameter α = 1. We obtained qualitatively similar results with different grid sizes and parameters (not shown).

2.4 | Molecular diversity indices

We calculated the number of unique haplotypes, the number of polymorphic sites, the nucleotide diversity and the mean number of pairwise distances in Arlequin v3.5.2.2 (Excoffier & Lischer, 2010). With the same software, we assessed departures from the mutation-drift equilibrium, by calculating Tajima's D (Tajima, 1989) and Fu's F_S (Fu, 1997). We ran 1,000 coalescent simulations to assess significance of the parameters. We calculated all indices and statistics for *C. gariepinus* sl. for the entire continent, for each ichthyofaunal province, and for the main clusters identified in the haplotype network (Table S2).

2.5 | Phylogenetic inference, molecular dating and ancestral area reconstruction

We performed phylogenetic analyses on the dataset of unique haplotypes, obtained by collapsing all 443 sequences in FaBox (Villesen, 2007). We performed maximum likelihood (ML) and Bayesian inference (BI) tree searches in PhyML v3.0 (Guindon et al., 2010) and MrBayes v3.2 (Ronquist et al., 2011) respectively. Using a Bayesian information criterion (BIC), we selected the best-fitting models of sequence evolution in the smart model selection module (Lefort, Longueville, & Gascuel, 2017) implemented in PhyML. We assessed nodal support for the ML tree with 1,000 bootstrap replicates. For BI, Metropolis-coupled Markov chain Monte Carlo (MCMC) simulations (two independent runs, 2 million generations, eight chains, 25% burn-in) provided the posterior trees and model parameters. Split deviation frequencies were <0.01, indicating that MrBayes runs were run long enough. We assessed stationarity of chains and convergence of parameter values in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer) prior to constructing a 50% majority rule consensus tree. The post-burn-in effective sample sizes (ESS) for all parameters were >200, indicating that sampled

parameter values accurately represented the posterior distribution (Kuhner, 2009).

We inferred a chronogram with BEAST v1.8.0 (Drummond & Rambaut, 2007), employing a Bayesian skyline tree prior and a strict clock model. We selected this model as we mainly dealt with intraspecific data and as preliminary analyses in TREE-PUZZLE v5.3 (Schmidt, Strimmer, Vingron, & von Haeseler, 2002) indicated that a clock-like evolution could not be rejected at a 0.05 significance level. We ran two independent MCMC chains for 25 million generations, and sampled model parameters every 1,000 generations. We used LogCombiner (BEAST package) to combine the two chains, after having discarded the first 20% of generations as burn-in. We assessed the stationarity and convergence of parameter values in Tracer v1.6. Pooled post-burn-in ESS were >200. We computed a maximum-clade-credibility tree in TreeAnnotator (BEAST package), which we visualized with FigTree v1.4.1 (Rambaut, 2014). We calculated divergence times as mean node heights of the 95% highest posterior density (HPD) intervals. To calculate absolute divergence times, we assumed substitution rates of 0.75 and 2.2% per MY, a range typically observed and applied for *cyt*b in fish (e.g. Bermingham, McCafferty, & Martin, 1997; Doadrio & Carmona, 2004; Zhao et al., 2009).

We performed ancestral area reconstructions in RASP v3.2 (Yu, Harris, Blair, & He, 2015) with both the S-DIVA (Yu, Harris, & He, 2010) and S-DEC model (Beaulieu, Tank, & Donoghue, 2013; Ree & Smith, 2008). To take topological uncertainty into account, we used 1,000 randomly sampled post-burn-in trees from both the MrBayes and the BEAST analyses as input. The five ichthyofaunal provinces from which fish were sampled were used as area definitions: East Coast, Congo, Nilo-Sudan, Zambezian and the Southern province. As samples from 'Lake Albert' originated both from the Lake proper as from above the escarpment, they were, for this analysis, included in the East Coast province. We defined an ancestral range to include no more than two adjacent areas, as this is also the current maximum observed range of haplotypes. We visualized the results of the RASP analyses on a map, taking the current distribution of terminal nodes into account.

2.6 | Microsatellite genotyping

We complemented *cyt*b sequencing by a genotyping approach. For this, we focussed on the south-eastern part of the Congo basin, where mitochondrial data revealed high genetic diversity. We genotyped 280 adult *C. gariepinus* from eight sites at seven loci, *Cga01*, *Cga02*, *Cga05*, *Cga09*, *Cga11* (Galbusera, Volckaert, Hellemans, & Ollevier, 1996), *Cba 19* and *Cba 20* (Yue, Kovacs, & Orban, 2003). We amplified all loci using the QIAGEN Multiplex PCR kit (QIAGEN), following the conditions in Table S3, and ran multiplex PCR products with an internal size standard (500LIZ, Applied Biosystems), on an ABI 3130 automated capillary DNA sequencer (Applied Biosystems). We conducted fragment analysis in GENEMAPPER v4.0 (Applied Biosystems) following the recommendations of Larmuseau, Raeymaekers, Hellemans, Van Houdt, and Volckaert (2010). We checked for the potential occurrence of null alleles and scoring errors, with MICRO-CHECKER v2.2.3 (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004) and DROPOUT v2.0 (McKelvey & Schwartz, 2005). Scoring errors were detected at loci *Cga09* and *Cba19*, which were excluded from further analyses. We ran 10%–30% of the samples twice for all markers to verify reproducibility.

We calculated the mean number of alleles per locus, the observed (H_O) and unbiased expected ($H_{E\ nb}$) heterozygosity, and the deviation from Hardy–Weinberg equilibrium (F_{IS}) in each population with GENETIX v4.05 (Belkhir, Borsa, Chikhi, Raufaste, & Bonhomme, 2004). We assessed allelic richness (corrected for sample size) using FSTAT v2.9.3.2 (Goudet, 2001) and tested linkage disequilibrium (LD) between the pairs of loci using GENEPOP v3.4 (Raymond & Rousset, 1995). We quantified population differentiation with F_{ST}, estimated as Θ (Weir & Cockerham, 1984), and R_{ST}, estimated as ρ (Slatkin, 1995). We calculated F_{ST}-linked pairwise genetic distances (Cavalli-Sforza & Edwards, 1967; D_{CE}) in GENETIX and R_{ST}-linked pairwise genetic distances (Goldstein, Linares, Cavalli-Sforza, & Feldman, 1995; $d\mu^2$) in SPAGeDI (Hardy & Vekemans, 2002). We tested F_{ST} and D_{CE} values for significance against 104 random permutations with a sequential Bonferroni test. We performed multidimensional scaling (MDS) in STATISTICA (Statsoft 2009), using the D_{CE} and $d\mu^2$ distances. We carried out a Bayesian clustering analysis in STRUCTURE v2.2 (Pritchard, Stephens, & Donnelly, 2000) using the no-admixture model without prior population information. We ran 10 replicate analyses (100,000 MCMC iterations, following a burn-in of 10,000) for K = 1–10 (number of clusters) and selected the K with the largest difference in log-likelihoods (ΔK) as the most likely number of clusters (Evanno, Regnaut, & Goudet, 2005). Finally, we performed Mantel tests (Smouse, Long, & Sokal, 1986) in GENETIX to analyse the effect of geographical on genetic distance, using both types of pairwise genetic distances.

3 | RESULTS

3.1 | Haplotype network

We identified seven clusters in the TCS network (Figure 2), which we labelled according the dominant origin of their haplotypes (Figure 1), obtaining NS (Nilo-Sudan), C (Congo), K (Kivu), GL (Great Lakes: Victoria, Tanganyika, Edward and Albert), S (Southern), T (Tana) and Z (Zambezi) clusters (Figure 2; Table S2). The K cluster occupied a central position in the network, but was closest to the genetically diverse C cluster (3 vs. 7–9 mutations respectively). Although most samples from the Congo province had C cluster haplotypes, specimens from the basin's northern, southern and eastern boundaries belonged to the NS, S and GL clusters respectively. A more complex pattern was observed in the Zambezi province. Here, all specimens from the western sector of the province, which includes the Kunene, the Okavango and the Upper Zambezi, carried S cluster haplotypes. Specimens from the province's eastern sector: the Lower Zambezi and Lake Malawi, all bore C cluster

haplotypes. In the central sector: the Middle Zambezi and the Limpopo basin, specimens mostly had Z cluster haplotypes. All specimens from the Southern province bore S cluster haplotypes except for some from coastal streams near the border with the Zambezi province that had Z cluster haplotypes. Specimens from the East Coast province had haplotypes belonging to the K, GL and T clusters. In the Nilo-Sudan province, all specimens bore NS haplotypes, except those from Lake Albert, which had GL cluster haplotypes.

As we noticed additional structuring in the NS, Z, S and C clusters, we divided them into two, three, four and six subclusters respectively (Table S2; Figure 2). Within the NS cluster, one subcluster consisted of sequences of *C. anguillaris* from West Africa, and *C. gariepinus* from the lower Nile and the Ubangi (Congo basin), whereas the other was restricted to *C. gariepinus* specimens from the lower Nile. Two of the three Z subclusters consisted of haplotypes that mostly occurred in sympatry in the central zone of the Zambezi province, although the range of one of them also extended into the Southern province. The third Z subcluster was restricted to populations from the Pongola River, which also harboured specimens with S cluster haplotypes. Three of the four S subclusters were mostly confined to populations from the Southern province. Specimens of the fourth S subcluster, which forms the link with the remainder of the network, stem from the Upper Lualaba and Upper Kasai (Congo basin), and from the Cunene, Okavango and Upper and Middle Zambezi (Zambezi province). Three of the six subclusters of the genetically diverse C cluster were restricted to the Congo basin. Here, haplotypes of one of these occurred throughout the basin whereas those of the other two were restricted to the Lufira River in the basin's south. Another C subcluster was shared between *C. gariepinus* from the Bangweulu-Mweru ecoregion (south-eastern Congo) and *Bathyclarias* from Lake Malawi. The last two C subclusters were restricted to specimens from Lake Malawi and the Lower Zambezi respectively.

3.2 | Genetic landscape shape interpolation

The genetic landscape analysis revealed two large zones of low genetic differentiation in *C. gariepinus* sl., one at the northern and one at the south-western part of the continent (Figure 3). We also observed low genetic differentiation along the borders of the Congo basin. Yet, sharp differentiation was present within the basin itself, especially in its northern and southern parts. Genetic differentiation was lower in the east, which is in line with the central position of Kivu specimens in the network. The East African Great Lakes (except Lake Malawi) also formed a region with low genetic differentiation. Sharp differentiation was present within the Zambezi province, reflecting the distinctness of its eastern, central and western sectors. On a smaller geographical scale, the analysis revealed the distinction between populations from the upper and the lower reaches of the easternmost rivers of the Southern province.

3.3 | Molecular diversity indices

Diversity indices are summarized in Table 1. All parameters show that the Southern ichthyofaunal province has the lowest genetic

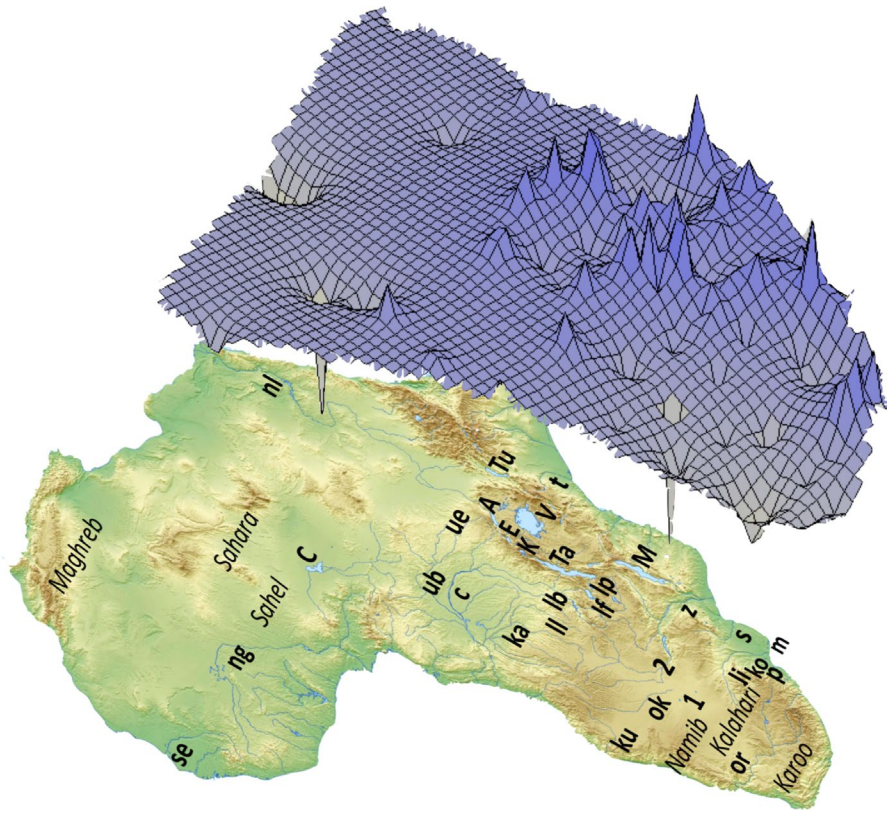

FIGURE 3 Results of the genetic landscape shape interpolation analysis of *C. gariepinus* sl. High values denote high and low values denote low genetic differentiation. Analyses were performed with distance weight parameter $\alpha = 1$ and visualized on a grid size of 1 × 1 degree. The topographical map of Africa was downloaded from https://mapswire.com. Localities mentioned in the text are indicated as: lakes: A Albert, C Chad, E Edward, K Kivu, M Malawi, Ta Tanganyika, Tu Turkana, V Victoria, rivers: c Congo, ka Kasai, ko Komati, ku Kunene, li Limpopo, lb Lualaba, lp Luapula (draining the Bangweulu-Mweru ecoregion), lf Lufira, ll Lulua, m Maputo, ng Niger, nl Nile, ok Okavango, or Orange, p Pongola, sa Save, se Senegal, t Tana, ue Uele, ub Ubangi, z Zambezi, and the 1 Makgadikgadi salt pans and 2 Victoria Falls [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Molecular diversity indices calculated on 418 *Clarias gariepinus sensu lato* haplotypes

Group	*N*/H	S	He (sd)	π (sd)	Δ (sd)	D	F_S
Africa	418/72	84	0.933 (0.005)	0.022(0.011)	11.093 (5.050)	−0.369	−15.747*
Provinces							
Congo	116/21	49	0.811 (0.028)	0.015 (0.008)	7.544 (3.547)	−0.562	0.425
East Coast	52/14	23	0.809 (0.038)	0.009 (0.005)	4.486 (2.245)	−0.382	−1.038
Nilo-Sudan	11/7	21	0.909 (0.066)	0.015 (0.009)	7.564 (3.826)	0.250	0.673
Southern	109/9	17	0.704 (0.023)	0.007 (0.004)	3.457 (1.779)	2.472	0.197
Zambezi	130/29	45	0.758 (0.037)	0.009 (0.005)	4.634 (2.288)	−1.350	−8.091*
Clusters							
C	111/20	34	0.787 (0.029)	0.010 (0.005)	4.874 (2.395)	−0.739	−1.823
NS	10/6	11	0.889 (0.075)	0.009 (0.006)	4.622 (2.477)	0.845	0.203
S	132/15	17	0.776 (0.020)	0.006 (0.003)	2.927 (1.545)	−0.165	−1.609
GL	55/9	9	0.725 (0.044)	0.004 (0.003)	2.044 (1.165)	0.108	−0.926
Z	104/15	16	0.584 (0.052)	0.003 (0.002)	1.568 (0.943)	−1.360	−7.593**

Group: clustered either by continent (Africa); ichthyofaunal province (according to Snoeks and Getahun (2013)) or TCS cluster; N/H, number of sequences and unique haplotypes; S, number of polymorphic sites; He, haplotype diversity and standard deviation (sd); π, nucleotide diversity and standard deviation (sd); Δ, mean number of pairwise differences and standard deviation (sd); D, Tajima's D and F_S, Fu's F_S, * and ** denote significance from zero at the 0.05 and the 0.005 levels. Population genetic parameters were estimated only for provinces/clusters with sample sizes ≥10

diversity, whereas the Congo and Nilo-Sudan provinces harbour the most diverse *C. gariepinus* assemblages. In none of the provinces did Tajima's D deviate significantly from zero, whereas Fu's F_S was significantly negative for the whole continent and for the Zambezi province. We also calculated these parameters for the seven TCS clusters with more than 10 haplotypes. This revealed high diversity in the C and NS clusters, and lower diversity in the Z and S clusters. The sample size of the NS cluster (*N* = 10) could, however, already have been too small to make reliable estimates. Tajima's D did not significantly deviate from zero in any of the clusters whereas Fu's F_S was negative for the Z and T clusters, indicating recent population expansion.

3.4 | Phylogenetic reconstruction, molecular dating and ancestral area reconstruction

PhyML (ML) and MrBayes (BI) yielded highly similar phylogenetic trees (Figure S4a,b), but the BEAST-derived tree deviated in topology (Figure S5). Although all trees contained the same major clades, we observed large differences in their branching order. As these exclusively concerned nodes with low statistical support, the biogeographical hypotheses presented here should be interpreted with care (Figure 4d,e). Using 2.2% as the rate of molecular evolution, the BEAST-derived approach revealed that the split between *C. gariepinus* and *C. ngamensis* occurred at 1.65 MYA (2.3–1.1), whereas a 0.75% rate dated the split at 4.84 MYA (3.2–6.7; Figure S5). The deepest diversification in *C. gariepinus* sl. took place 0.85 MYA (1.2–0.6, 2.2%) or 2.5 MYA (1.7–2.5, 0.75%).

We used the MrBayes- and BEAST-derived trees as input files for RASP. The results for the S-DIVA (Figure 4a,b) and the S-DEC (Figure S6a,b) models were, overall, highly similar. However, the latter suggested more alternative scenarios and did not resolve the ancestral area for several nodes of the MrBayes tree. As this suggests that the S-DEC model could be less suitable to explain our data, we choose to focus on the area reconstructions provided by S-DIVA. This model suggested, for both input trees, that the common ancestor of *C. gariepinus* sl. and *C. ngamensis* occurred both in the current Congo and East Coast provinces, whereas it placed the origin of *C. ngamensis* in the Congo (less support for a Congo-Zambezi distribution). The origin of *C. gariepinus* sl. was, for both input trees, placed in the East Coast, although the BEAST-derived trees also showed minor support for a joined East Coast/Zambezi origin (Figure 4b,e).

The MrBayes-based RASP analysis (Figure 4a) revealed a pattern where a first diversification occurred in the Great Lakes region, with a first lineage colonizing Lake Kivu and spreading further towards the Congo and the eastern Zambezi. The deepest node explaining the spread of this lineage towards the latter two provinces was unresolved, possibly because we maximally allowed for distributions of two provinces. As the basal clades in this lineage are restricted to regions near the boundary of the Congo, the East Coast and the Zambezi provinces, initial diversification might have occurred there. Later on this lineage spread further within the Congo and Lower Zambezi basins. As several internal nodes in this lineage were partially or fully resolved as having Congo–Zambezi distributions, diversification most likely happened through vicariance events. A second lineage also originated in the East Coast. This contained a basal clade that remained restricted to the Great Lakes, and two lineages that spread north and southwards. The northern colonised the Nilo-Sudanic province, from where it also entered Congo. The southern lineage spread throughout the Zambezi province, from where it invaded the south of the Congo basin and the Southern province.

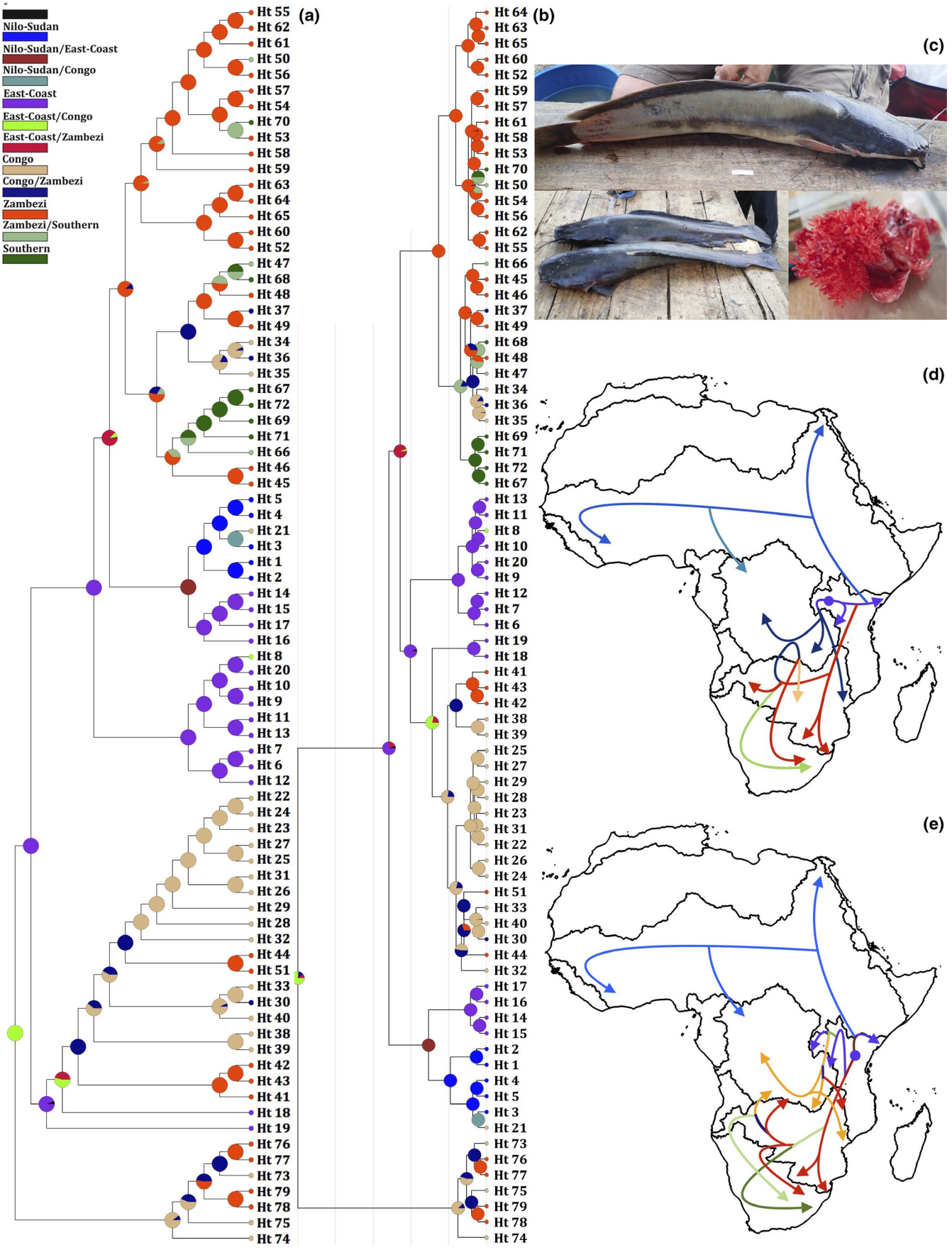
(a)
(b)
(c)
(d)
(e)
Nilo-Sudan
Nilo-Sudan/East-Coast
Nilo-Sudan/Congo
East-Coast
East-Coast/Congo
East-Coast/Zambezi
Congo
Congo/Zambezi
Zambezi
Zambezi/Southern
Southern
Ht 55 Ht 62 Ht 61 Ht 50 Ht 56 Ht 57 Ht 54 Ht 70 Ht 53 Ht 58 Ht 59 Ht 63 Ht 64 Ht 65 Ht 60 Ht 52 Ht 47 Ht 68 Ht 48 Ht 37 Ht 49 Ht 34 Ht 36 Ht 35 Ht 67 Ht 72 Ht 69 Ht 71 Ht 66 Ht 46 Ht 45 Ht 5 Ht 4 Ht 21 Ht 3 Ht 1 Ht 2 Ht 14 Ht 15 Ht 17 Ht 16 Ht 8 Ht 20 Ht 10 Ht 9 Ht 11 Ht 13 Ht 7 Ht 6 Ht 12 Ht 22 Ht 24 Ht 23 Ht 27 Ht 25 Ht 31 Ht 26 Ht 29 Ht 28 Ht 32 Ht 44 Ht 51 Ht 33 Ht 30 Ht 40 Ht 38 Ht 39 Ht 42 Ht 43 Ht 41 Ht 18 Ht 19 Ht 76 Ht 77 Ht 73 Ht 79 Ht 78 Ht 75 Ht 74
Ht 64 Ht 63 Ht 65 Ht 60 Ht 52 Ht 59 Ht 57 Ht 61 Ht 58 Ht 53 Ht 70 Ht 50 Ht 54 Ht 56 Ht 62 Ht 55 Ht 66 Ht 45 Ht 46 Ht 37 Ht 49 Ht 68 Ht 48 Ht 47 Ht 34 Ht 36 Ht 35 Ht 69 Ht 71 Ht 72 Ht 67 Ht 13 Ht 11 Ht 8 Ht 10 Ht 20 Ht 9 Ht 12 Ht 7 Ht 6 Ht 19 Ht 18 Ht 41 Ht 43 Ht 42 Ht 38 Ht 39 Ht 25 Ht 27 Ht 29 Ht 28 Ht 23 Ht 31 Ht 22 Ht 26 Ht 24 Ht 51 Ht 33 Ht 40 Ht 30 Ht 44 Ht 32 Ht 17 Ht 16 Ht 14 Ht 15 Ht 2 Ht 1 Ht 4 Ht 5 Ht 3 Ht 21 Ht 73 Ht 76 Ht 77 Ht 75 Ht 79 Ht 78 Ht 74

FIGURE 4 Reconstruction of the spread of *C. gariepinus* sl. throughout Africa. Ancestral state reconstruction using RASP with the S-DIVA model on (a) a Bayesian inferred phylogeny reconstruction constructed using MrBayes and (b) a BEAST-constructed chronogram. These reconstructions are visualized on a map of the continent with (d) the MrBayes- and (e) the BEAST-derived scenario. (c) *C. gariepinus* from Lake Edward (top), *Bathyclarias* spp. from Lake Malawi (bottom left) and the suprabranchial organ of *C. gariepinus* (bottom right; pictures, MVS and N. Vranken). Colours of nodes (a,b) and arrows (d,e) agree with the one or two ichthyofaunal provinces where the diversification took place as indicated in the legend top left. Codes for haplotypes as showed in Tables S1 and S2 [Colour figure can be viewed at wileyonlinelibrary.com]

The BEAST-derived approach (Figure 4b) supported a scenario where an East Coast population split into a northern line, colonizing the Nilo-Sudanic province and the Tana River, and a southern line. A vicariance event separated the latter into a Zambezian and a Great Lakes lineage. Of these, the former spread throughout the Zambezi province and invaded the south of the Congo basin and the Southern province. The latter spread throughout the Great Lakes and invaded, via Lake Kivu, the Congo and the eastern Zambezi. In contrast to the MrBayes approach, the Congo receives the most support as the origin of the joint Congo-eastern Zambezi lineage.

3.5 | Microsatellite genotyping

We investigated the population structure of *C. gariepinus* in the Congo basin using microsatellites (Dryad https://doi.org/10.5061/dryad.f1vhhmgt4). Mean allelic richness per population varied between 4.3 and 7.6 (mean = 5.5) and mean expected heterozygosity ranged between 0.5 and 0.7 (mean = 0.6). Five populations had a significant departure from Hardy–Weinberg equilibrium and pairwise comparisons between loci revealed no significant linkage disequilibrium (Table 2). Pairwise F_{ST} and R_{ST} values are summarized in Table 3. The D_{CE}-based MDS plot (Figure 3) revealed that the three populations from the main course of the Congo (KIN, KIS and DBL) clustered together. The same held for the populations from the Lulua and the Upper Lufira (KAD, LUF). A final cluster contained the populations from the Middle Lufira and the Luapula basin (KYU, KAF, LUT). Both MDS axes had stress values below 0.20, indicating that the inter-site relationships can be interpreted (Clarke, 1993). STRUCTURE revealed a separation into the same hypothetical clusters. Although this analysis also assigned specimens from Mwaba (LUT) to the third cluster, total population assignment was only 68% (Figure 5; Table 2). Finally, Mantel tests revealed no isolation by distance, neither with D_{CE} ($r = .170$, $p > .05$) nor with $d\mu^2$ ($r = .013$, $p > .05$). All populations of microsatellite cluster I carried mtDNA haplotypes of subcluster C6. For cluster II, one population bore a S1 and the other contained specimens with both C2 and C6 haplotypes. For cluster III, two populations bore C4 and the third the closely related C5 haplotypes.

4 | DISCUSSION

4.1 | Validating biogeographical entities

We hypothesized that the genetic structure of *C. gariepinus* sl. can be used to validate biogeographical entities such as proposed ichthyofaunal provinces. Although strong geographical patterns were evident from the haplotype network and the genetic landscape shape analysis (Figures 1-3), these did not fully match with the ichthyofaunal provinces. Most samples from the Nilo-Sudanic province grouped into a single cluster, confirming the homogeneity found in *C. gariepinus* and *C. anguillaris* of the region (Rognon et al., 1998), and that of their monogenean parasites (Barson, Přikrylová, Vanhove, & Huyse, 2010). This can be explained by previous connections between the region's main rivers during wetter phases of the late Quaternary (Drake & Bristow, 2006). The NS cluster also included the sample from the Ubangi (Congo basin), which supports the treatment of the Ubangi basin (save the Uele) as a distinct freshwater ecoregion: the Sudanic Congo. This ecoregion harbours many aquatic species of Nilo-Sudanic origin (Peck & Thieme, 2005) and is considered transitional between the Congo and the Nilo-Sudanic provinces (Thieme et al., 2005). Samples from Lake Albert did not cluster with other Nilo-Sudanic samples, but rather with those from the Great Lakes. This is remarkable as Snoeks and Getahun (2013) assigned the lake to the Nilo-Sudanic province, based on its ichthyofaunal similarity with the Nile.

Populations from the East Coast province were highly differentiated as samples from three distinct locations, Lake Kivu, the Tana River and lakes Edward and Victoria, occurred in different clusters, confirming Giddelo et al. (2002). Samples from Lake Kivu were ancestral to the C lineage, which spread throughout the Congo province. This seems in line with current drainage patterns as Lake Kivu drains, via Lake Tanganyika, into the Congo. However, Lake Kivu was not assigned to this province as its original fauna experienced extinction phases caused by the basin's turbulent tectonic history (Snoeks, De Vos, & Thys van den Audenaerde, 1997). We hypothesize that, due to its high physiological tolerance, *C. gariepinus* might have survived these adverse conditions. Hence, the current population might be one of the few surviving relicts of the original Kivu fauna. Additionally, the samples of *C. gariepinus* from Lake Tanganyika bore haplotypes that were identical to those found in Lake Victoria. An affinity between Tanganyikan and Victorian faunas is also known for the lakes' cichlid species flocks. Here, the ancestors of the haplochromines of lakes Kivu, Edward and Victoria can be traced to Lake Tanganyika (Verheyen, Salzburger, Snoeks, & Meyer, 2003). The distinctness of the samples from Lake Kivu suggests a different scenario for *C. gariepinus* and we speculate that the Tanganyika population of *C. gariepinus* consists of relative newcomers of East African origin. A similar scenario was put forward for representatives of *Oreochromis* Günther, 1889. The endemic *O. tanganicae* (Günther, 1894) and the widespread *O. niloticus* (Linnaeus, 1758) are both relatively new additions to the Tanganyika

TABLE 2 Summary of the microsatellite genotyping analysis of 280 *C. gariepinus* samples originating from eight locations in the Congo basin using five microsatellite markers

Site	Code	Location	River	Basin	N	E	N	AR	$H_{E\,nb}$	H_O	F_{IS}	I	II	III
1	KIN	Kinshasa	Lower Congo	Congo	−04°18′	15°20′	20	4.523	0.5382	0.641	**−0.196**	**97.3**	1.2	1.6
2	KIS	Kisangani	Middle Congo	Congo	00°31′	25°11′	38	6.552	0.6861	0.675	0.016	**94.6**	1.2	4.3
3	DBL	Bukama	Upper Congo	Congo	−09°11′	25°51′	47	4.920	0.5739	0.515	**0.103**	**97.8**	1	1.3
4	KAD	Dianda	Kando	Lulua	−10°45′	25°50′	26	5.625	0.6883	0.599	**0.133**	2.6	**89.8**	16.6
5	LUF	Kapolowe	Upper Lufira	Lufira	−11°21′	26°46′	47	4.297	0.5243	0.353	**0.328**	1.4	**93.0**	5.5
6	KYU	Kyubo	Middle Lufira	Lufira	−09°31′	27°02′	48	4.601	0.4754	0.447	0.061	2.9	4.3	**92.8**
7	KAF	Lubumbasi	Kafubu	Luapula	−11°42′	27°29′	25	5.550	0.5763	0.392	**0.325**	6.5	11.4	**82.2**
8	LUT	Mwaba	Lutshipuka	Luapula	−10°14′	28°20′	29	7.576	0.6953	0.691	0.006	22.0	10.3	**67.7**

A summary of the material includes the site number, location, river and main river basin and the number of samples used. Estimates of genetic diversity include: AR, allelic richness; $H_{E\,nb}$, unbiased expected heterozygosity; H_O, observed heterozygosity; F_{IS}, deviation from the Hardy–Weinberg equilibrium, with statistically significant F_{IS} values listed in bold. I, II and III denote the proportion (in %) of the populations assigned to the three hypothetical clusters identified by STRUCTURE, with the highest assignment value for each population listed in bold.

cichlid assemblage (Klett & Meyer, 2002). *Oreochromis* also consists of highly tolerant, generalist species. Hence, it is not unlikely that their demographics mirror that of *C. gariepinus*. As *Oreochromis* has its centre of diversity in the East Coast, the Tanganyika representatives of *Oreochromis* and *C. gariepinus* might have used a similar colonization route. Finally, we cannot exclude that, next to the GL haplotypes, ancestral Kivu haplotypes might have been present in *C. gariepinus* from Lake Tanganyika.

Both microsatellites and mtDNA genotypes showed a large degree of homogeneity in the centre of the Congo province, which did not extend to the basin's boundaries. Here, the distribution of mitochondrial lineages reflected previous connections with watercourses that now drain to the Nile, Lake Victoria (see above) and the Zambezi. This pattern reflects multiple invasions from neighbouring basins, similar to what was seen in representatives of *Mastacembelus* (Day et al., 2017). In the south of the basin, two populations from the Upper (Lualaba) and Middle Congo (Kasai) shared S cluster haplotypes with *C. gariepinus* from the western Zambezi. This confirms Skelton (1994), who noticed that present-day watersheds divide rather than circumscribe the northern boundaries of the Zambezian ichthyofauna. Parasitological data derived from monogeneans of cichlids (Vanhove et al., 2013) and clariids (Barson et al., 2010) also support this hypothesis. The distribution of S cluster lineages is also reminiscent of that of serranochromine cichlids. These constitute a prominent part of the Upper Zambezi ichthyofauna and have, aided by recent river capture events, spread to neighbouring basins (Joyce et al., 2005; Katongo, Koblmüller, Duftner, Mumba, & Sturmbauer, 2007).

The rivers that drain the southern plateaus of the Congo are characterized by waterfalls that isolate its fish fauna from that of the rest of the basin. This explains why the region's *C. gariepinus* populations often carried C cluster haplotypes that deviated strongly from those found in the basin's centre. This also held for populations from the Bangweulu-Mweru ecoregion, the south-easternmost part of the basin. This confirmed the distinct, yet Congolese origin of the region's ichthyofauna (Van Steenberge, Vreven, & Snoeks, 2014), for which it was previously assumed that it derived from the Zambezi (Scott, 2005). Microsatellite analyses grouped samples from this ecoregion (Figure 5, KAF, LUT) with those of the Middle Lufira (KYU) although they also partially assigned specimens from the most downstream part of the ecoregion (LUT) to the widespread cluster I. A similar pattern was found in tigerfish, *Hydrocynus*. Here, the ecoregion was home to a unique mitochondrial lineage, which co-occurred with a more widespread strain (Goodier et al., 2011). Interestingly, some representatives of *Bathyclarias* from Lake Malawi shared haplotypes with the Bangweulu-Mweru population of *C. gariepinus*. Other specimens of *Bathyclarias* had a closely related haplotype, which they shared with the local *C. gariepinus* from Lake Malawi. This suggests a recent connection between the basin of Lake Malawi and this part of the Congo basin. The close affinity between Malawi and Luapula populations of the basal haplochromine cichlid *Pseudocrenilabrus philander* (Weber, 1897; Egger et al., 2015) points in the same direction. Other specimens of Lake

TABLE 3 Pairwise genetic differentiation of *C. gariepinus* in the Congo basin

/	1	2	3	4	5	6	7	8
1	–	**0.038**	0.125	0.409	**0.735**	**0.351**	0.439	**0.235**
2	0.017	–	0.092	0.408	**0.713**	**0.281**	**0.425**	**0.176**
3	**0.108**	**0.085**	–	**0.303**	**0.420**	**0.401**	**0.311**	**0.240**
4	**0.307**	**0.243**	0.489	–	**0.549**	**0.308**	0.389	**0.183**
5	**0.437**	**0.366**	**0.705**	**0.198**	–	**0.424**	**0.352**	**0.339**
6	0.536	**0.516**	0.489	0.299	0.233	–	0.257	**0.212**
7	**0.272**	**0.213**	0.498	**0.203**	0.599	**0.113**	–	**0.163**
8	0.418	**0.408**	0.382	0.323	0.292	**0.323**	0.245	–

Pairwise F_{ST} (above diagonal) and R_{ST} (below diagonal) values of the *C. gariepinus* from the Congo basin samples based on five microsatellite markers. Statistically significant values (after Bonferroni correction) are listed in bold. See Table 2 for sample codes.

Malawi *Bathyclarias* bore haplotypes that were only distantly related to this lineage, pointing at an earlier colonization by the parent species of *Bathyclarias*: *C. gariepinus*. Other phylogeographical studies of aquatic taxa indicated several cases of contact between the basins of the Congo and of Lake Malawi. Lake Malawi's endemic species of *Bellamya* Jousseaume, 1886 (Gastropoda, Viviparidae) were shown to be sister to an endemic Congo lineage, pointing at a very old connection (Schultheiss et al., 2014). Some of the Lake's three species of *Opsaridium* Peters, 1854 (Cyprinidae) might also have their origin in the Congo basin (Sungani et al., 2017). Finally, a parallel can be drawn between the evolutionary mechanisms that gave rise to the radiations of *Bathyclarias* and haplochromine cichlids in Lake Malawi. In both cases, the presumed parent species of these radiations, *C. gariepinus* and *Astatotilapia calliptera* (Günther,

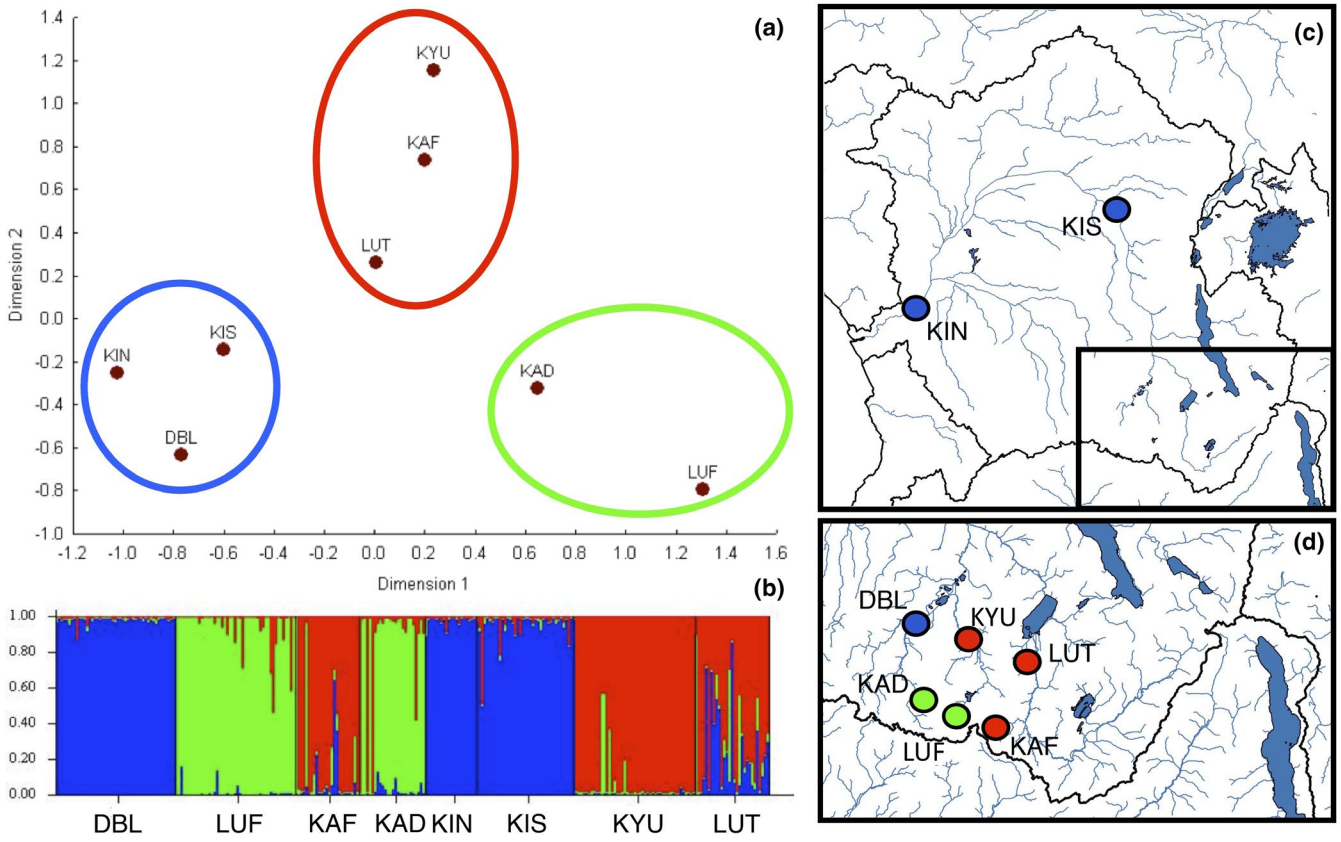

FIGURE 5 Analyses of the microsatellite data on 280 specimens of *C. gariepinus* belonging to eight populations. (a) MDS plot of pairwise genetic distances based using D_{CE}, (b) clustering analysis conducted in STRUCTURE 2.2 with K = 3 and with each vertical bar representing the estimated proportion of membership to the three clusters, (c,d) map of the Congo ichthyofaunal province and of its south-easternmost part (d) indicating the sample origins. Sample codes as in Table 2, colours denote the three microsatellite-based clusters derived using STRUCTURE with I: blue, II: green and III: red [Colour figure can be viewed at wileyonlinelibrary.com]

1894), respectively, occur in sympatry with the 'offspring species'. Phylogenies of these two radiations contain members of the parent species and the members of the two species flocks form clades within the diversity of the parent species (Agnèse & Teugels, 2001; Malinsky et al., 2018). Strangely, a widespread haplotype from the central part of the basin was also found in the Upper Lufira, where it occurred in sympatry with a highly divergent C2 haplotype. Although it cannot be excluded that the former stems from an anthropogenic introduction, this seems unlikely as microsatellite data did not reveal any evidence of introgression. The deviant haplotype, which is most closely related to one found in Lake Malawi populations, reflects the relict fauna of the Lufira (Van Steenberge, Gajdzik, Chilala, Snoeks, & Vreven, 2014).

The complex geological and hydrological history of southern Africa, that is the Zambezi and Southern ichthyofaunal provinces, is mirrored by the distribution of mitochondrial haplotypes in its populations of *C. gariepinus*. We delineated three zones: an eastern, central and western that, with the exception of a few boundary locations, bore mitochondrial haplotypes of the C, Z and S clusters respectively. The eastern zone contained the Lower Zambezi and Lake Malawi. Populations of this zone were related to those from the Congo basin (see above). The central zone included the Limpopo, Komati and Maputo Rivers as well as the middle section of the Zambezi. Hydrological links between these water bodies have already been suggested to explain hybridization events in monogenean parasites of clariids (Barson et al., 2010). The western zone, which is known to have a different ichthyofauna than the rest of the province (Skelton, 1994), included the western part of the Zambezi, and the entire Southern province.

Originally, both the western and the central part of the Zambezi system drained to the Indian Ocean via the Limpopo (Stankiewicz & de Wit, 2006). Uplift severed this connection, and these rivers became connected with the Middle and Lower Zambezi, creating the Victoria Falls. Yet, it has been hypothesized that during the mid-Pleistocene, this connection was disrupted again, giving rise to a vast lake, Palaeo Makgadikgadi (Joyce et al., 2005). As this lake underwent several cycles of expansion and retraction, its formation has been difficult to date geologically (Derricourt, 1976). Our molecular clocks suggest that the separation of the *C. gariepinus* populations from the Limpopo and the Upper Zambezi occurred at about 400 KYA (600–300 KYA), when using a 2.2% rate and 1.17 MYA (1.76–0.88 MYA), when using a 0.75% rate of molecular evolution. The estimates calculated with the faster rate correspond with the original formation of the Palaeolake (Moore & Larkin, 2001).

Ancestral area reconstructions revealed that the populations of *C. gariepinus* currently inhabiting the western zone of southern Africa spread from a region that at least encompassed part of the current Zambezi ichthyofaunal province. From there, *C. gariepinus* entered the Southern province at least twice. This might point towards a larger south-west draining Trans-Tswana system (McCarthy, 1983), that might have drained a large part of south-western Africa to the Atlantic Ocean via the current day Orange River. The system might have been connected to the basin of Palaeolake Makgadikgadi. Alternatively, even in the absence of proper river connections, the landscape matrix of this part of the continent would, at periods of higher rainfall, have allowed *C. gariepinus* to disperse from the Zambezi into the Orange basin. This scenario fits with climate reconstructions of southern Africa. Alternating deposits of dune sands in the Kalahari basin revealed cycling between dry and wet conditions in parallel with glacial cycles at higher latitudes (Munyikwa, Van Den Haute, Vandenberghe, & De Corte, 2000). This scenario also explains why the ecologically tolerant *C. gariepinus* was able to disperse southward, whereas more specialized taxa, such as species belonging to *Hydrocynus*, *Synodontis* or *Mastacembelus*, failed to do so.

The eastern border between the Southern and the Zambezi ichthyofaunal provinces is not well defined, and many coastal streams have a fish fauna that is a mixture of both provinces (Skelton, 1994). The genetic structure of the *C. gariepinus* populations of these basins revealed how these mixed faunas emerged. The lower reaches are dominated by typical 'central Zambezian' Z cluster haplotypes, whereas in the upper reaches, specimens are found with haplotypes related to those from the rest of the Southern province. This suggests multiple colonizations of these rivers by river capture events in the highlands, in parallel with invasions via the coastal plain. Of interest is that the distribution of S- and Z-type haplotypes matches the distinction between two freshwater ecoregions, whose boundaries do not follow ichthyofaunal provinces: the 'Zambezian lowveld', downstream, and the 'southern temperate highveld', upstream (Thieme et al., 2005). The separation between the up and downstream populations of *C. gariepinus* highlights the importance of waterfalls in maintaining the isolation of the ichthyofaunal communities of both ecoregions.

4.2 | Reconstructing landscape evolution and climatic events

Clarias gariepinus is unique because of its almost continuous distribution in Africa. Other widespread African fish species, with distribution ranges that span several provinces, include *Heterobranchus longifilis* Valenciennes, 1840, *Hydrocynus vittatus* (Castelnau, 1861) and *Mormyrops anguilloides* (L., 1758; Mormyridae). These species all have, in contrast to *C. gariepinus*, large gaps in their distributions. Such discontinuous ranges have been attributed to local extinctions that erased patterns laid down during the initial spread of a lineage (Cotteril & de Wit, 2011). We hypothesize that, due to its ecological tolerance, *C. gariepinus* survived the unfavourable conditions that caused other taxa to go extinct. The same trait might have prevented (allopatric) speciation, allowing multiple colonization events to be conserved in the genome of a single population in the form of distantly related haplotypes.

The distribution of haplotypes in widespread species can be used to document temporal and spatial changes in climate. Ancestral area reconstructions revealed that the deepest splits within *C.*

gariepinus, as well as the origin of the species itself, are found in the region of the East African Great Lakes. A vicariance event between an East Coast and a Congolese lineage seemed to be at the origin of *C. gariepinus* and *C. ngamensis*. This suggests an important role for rifting, which affected the climate in Africa (Maslin et al., 2014). When using the 2.2% estimate of molecular evolution, this split was dated at about 1.65 MYA, a time when both aridity and climatic variability across Africa increased (dated at 1.9–1.5 MYA by Trauth, Larrasoaña, & Mudelsee, 2009). Additionally, the first diversification event within *C. gariepinus*, and hence the spread of the lineage, coincided with the Early Middle Pleistocene transition, representing the first truly dry period in East Africa (1 MYA, Maslin et al., 2014). The more severe impact of climate change in East Africa compared to the Congo basin might explain why *C. gariepinus* evolved a higher ecological tolerance than its sister species *C. ngamensis*. The onset of African aridification, which happened in a stepwise way, already started earlier in the mid-Pliocene (deMenocal, 2014). Hence, also slower estimates of molecular evolution support this evolutionary scenario.

Pleistocene fluctuations in climate also shaped the distribution of terrestrial taxa. The phylogeography of large savanna mammals typically reveals three major clades that match with refugia in the continent's western, eastern and southern savanna zones (Hewitt, 2004). These zones are mirrored in the phylogeographical patterns of *C. gariepinus*. The MrBayes-derived phylogeography (Figure 4a) of *C. gariepinus* can be interpreted as branching in a forest and a savanna clade, with the latter consisting of subclades occurring in these three savanna zones. The BEAST-derived phylogeography reveals a similar pattern, although here, the 'forest' clade clusters within the eastern group (Figure 4b). The genetic landscape analysis revealed two large zones of low genetic differentiation: one at the northern and one at the south-western part of the continent. These regions currently contain some of the driest regions of Africa: the Sahara and the Sahel in the north, and the Kalahari, Namib and Karoo in the south. At first sight, this is counterintuitive as one would expect these regions to represent barriers to fish migration. Yet, the occurrence of very closely related haplotypes within and close to these current arid zones reflects the presence of a permeable landscape matrix for *C. gariepinus* in the relatively recent past. These conditions would have been present during the African wet period that followed the last glacial maximum. In the north of the continent, the now extinct megalakes of the Sahara, of which Palaeolake Megachad was the largest, provided the opportunity for faunal connections between Nile, Niger and Ubangi (Drake & Bristow, 2006). In the south, Palaeolake Makgadikgadi might have played a similar role (Joyce et al., 2005). Interestingly, this landscape might not have been equally permeable for other African freshwater fish. In the Nilo-Sudanic province, conspecific populations of *Synodontis* were significantly divergent in mitochondrial markers, whereas this clade failed to spread to the Southern ichthyofaunal province (Day et al., 2013).

The ecological characteristics and distribution of *C. gariepinus* make it an ideal choice to study successive stages of landscape evolution. The sole caveat, however, is that since the species became well-established in aquaculture, human-induced translocations have left signatures in its genome. Due to translocation, the species has become part of the non-indigenous fish community in many tropical and subtropical regions (Radhakrishnan, Lan, Zhao, Qing, & Huang, 2011; Vitule, Umbria, & Aranha, 2006). For the current dataset, this is, of course, the case for the feral population from the Western Cape (Cambray, 2003). As the distribution of haplotypes in *C. gariepinus* mirrors faunal or climatic regions, it can be assumed that populations are locally adapted. Translocations and introductions of non-native strains therefore represent threats to this important fishery resource. Given the increased farming of *Clarias*, in which often only a few commercial strains are used, local stocks have to be promoted (Roodt-Wilding, Swart, & Impson, 2010).

ACKNOWLEDGEMENTS

A.A. received a postdoctoral research fellowship from the KU Leuven and the Research Foundation - Flanders (FWO). M.V.S. and M.P.M.V. received PhD fellowships from FWO. MVS received travel grants of the FWO and the Leopold III funds for Nature Exploration and Conservation. A.C.M. received support from the Belgian Development Agency and the BTCCTB project PRODEPAAK. We sincerely thank the following colleagues for sample collection: J.-F. Agnèse, B. Benade, R. Bills, L. and G. Chapman, P. de Villiers, the late L. De Vos, J. Engelbrecht, P. Fouche, P. Galbusera, K. Goudswaard, D. Impson, R. Karssing, J. Munyandorero, A. Piers, C. Maguswi, J. Southey, the late G. Teugels, K. Van Lerberghe and the Department of Fisheries and Agriculture of Zambia. We benefited from insights from J.-F. Agnèse, T. Huyse, J. Maley, P. Skelton, J. Snoeks, K. Van Doninck and C. Sturmbauer. This study was performed on historical collections and on samples collected by researchers from certified institutes of the countries where collections were made. All material was collected prior to the implementation of the Nagoya protocol for access and benefit sharing.

DATA AVAILABILITY STATEMENT

All sequences are deposited in GenBank under accession numbers (MN941435-185 and AF126823-24; AF235922-28; AF235933; AF475153; DQ646360-72). Microsatellite data are deposited in Dryad under https://doi.org/10.5061/dryad.f1vhhmgt4.

ORCID

Maarten W. Van Steenberge https://orcid.org/0000-0002-6964-9014
Maarten P. M. Vanhove https://orcid.org/0000-0003-3100-7566
Maarten H. D. Larmuseau https://orcid.org/0000-0002-5974-7235
Jean-Claude Micha https://orcid.org/0000-0001-7973-3634
Stephan Koblmüller https://orcid.org/0000-0002-1024-3220
Rouvay Roodt-Wilding https://orcid.org/0000-0002-3528-7932
Filip A. M. Volckaert https://orcid.org/0000-0003-4342-4295

REFERENCES

Agnèse, L. F., & Teugels, G. G. (2001). The *Bathyclarias-Clarias* species flock. A new model to understand rapid speciation in African Great lakes. *Comptes Rendus De L'académie Des Sciences - Series III - Sciences De La Vie*, *324*, 683–688. https://doi.org/10.1016/s0764-4469(01)01348-8

Arndt, A., Van Neer, W., Hellemans, B., Robben, J., Volckaert, F., & Waelkens, M. (2003). Roman trade relationships at Sagalassos (Turkey) elucidated by ancient DNA of fish remains. *Journal of Archeological Science*, *30*, 1095–1105. https://doi.org/10.1016/s0305-4403(02)00204-2

Barson, M., Přikrylová, I., Vanhove, M. P. M., & Huyse, T. (2010). Parasite hybridization in African *Macrogyrodactylus* spp. (Monogenea, Platyhelminthes) signals historical host distribution. *Parasitology*, *137*, 1585–1595. https://doi.org/10.1017/s0031182010000302

Beaulieu, J. M., Tank, D. C., & Donoghue, M. J. (2013). A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. *BMC Evolutionary Biology*, *13*, 80. https://doi.org/10.1186/1471-2148-13-80

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (2004). *GENETIX 4.05*, *Logiciel Sous WindowsTM Pour la Génétique Des Populations*. Montpellier France: Université de Montpellier II.

Bermingham, E. S., McCafferty, A., & Martin, P. (1997). Fish biogeography and molecular clocks: Perspectives from the Panamanian Isthmus. In T. Kocher, & C. Stepien (Eds.), *Molecular Systematics of Fishes* (pp. 113–126). New York: Academic Press.

Burridge, C. P., Craw, D., Jack, D. C., King, T. M., & Waters, J. M. (2008). Does fish ecology predict dispersal across a river drainage divide? *Evolution*, *62*, 1484–1499. https://doi.org/10.1111/j.1558-5646.2008.00377.x

Cambray, J. A. (2003). The need for research and monitoring on the impacts of translocated sharptooth catfish, *Clarias gariepinus*, in South Africa. *African Journal of Aquatic Science*, *28*, 191–195. https://doi.org/10.2989/16085910309503786

Cavalli-Sforza, L. L., & Edwards, A. W. F. (1967). Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics*, *19*, 233–257.

Clarke, K. R. (1993). Nonparametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, *18*, 117–143. https://doi.org/10.1111/j.1442-9993.1993.tb00438.x

Cotterill, F. P. D., & de Wit, M. J. (2011). Geoecodynamics and the Kalahari epeirogeny: Linking its genomic record, tree of life and palimpsest into a unified narrative of landscape evolution. *South African Journal of Geology*, *114*, 489–514. https://doi.org/10.2113/gssajg.114.3-4.489

Cotterill, W., & Goodier, S. (2008). DNA clocks for dating landforms. *Quest*, *4*, 19–21.

Day, J. J., Fages, A., Brown, K. J., Vreven, E. J., Stiassny, M. L. J., Bills, R., … Rüber, L. (2017). Multiple independent colonizations into the Congo Basin during the continental radiation of African *Mastacembelus* spiny eels. *Journal of Biogeography*, *44*, 2308–2318. https://doi.org/10.1111/jbi.13037

Day, J. J., Peart, C. R., Brown, K. J., Friel, J. P., Bills, R., & Moritz, T. (2013). Continental diversification of an African Catfish Radiation (Mochokidae: *Synodontis*). *Systematic Biology*, *62*, 351–365. https://doi.org/10.1093/sysbio/syt001

deMenocal, P. B. (2014). African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planerary Science Letters*, *220*, 3–24. https://doi.org/10.1016/s0012-821x(04)00003-2

Derricourt, R. M. (1976). Regression rate of the Victoria Falls and the Bakota Gorge. *Nature*, *264*, 23–25. https://doi.org/10.1038/264023a0

Doadrio, I., & Carmona, J. A. (2004). Phylogenetic relationships and biogeography of the genus *Chondrostoma* inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, *33*, 802–815. https://doi.org/10.1016/j.ympev.2004.07.008

Drake, N., & Bristow, C. (2006). Shorelines in the Sahara: Geomorphological evidence for an enhanced monsoon from paleaolake Megachad. *The Holocene*, *16*, 901–911. https://doi.org/10.1191/0959683606hol981rr

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

Egger, B., Klaefiger, Y., Indermauer, A., Koblmüller, S., Theis, A., Egger, S., … Salzburger, W. (2015). Phylogeographic and phenotypic assessment of a basal haplochromine cichlid fish from Lake Chala, Zambia. *Hydrobiologia*, *748*, 171–184. https://doi.org/10.1007/s10750-014-1919-0

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*, 2611–2620. https://doi.org/10.1111/j.1365-294x.2005.02553.x

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analysis under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fairhead, J. D. (1988). Mesozoic plate tectonic reconstructions of the central South Atlantic Ocean: The role of the West and Central African Rift system. *Tectonophysics*, *155*, 181–191. https://doi.org/10.1016/0040-1951(88)90265-x

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*, 915–925.

Galbusera, P., Volckaert, F. A. M., Hellemans, B., & Ollevier, F. (1996). Isolation and characterization of microsatellite markers in the African catfish *Clarias gariepinus* (Burchell 1822). *Molecular Ecology*, *5*, 703–705. https://doi.org/10.1111/j.1365-294x.1996.tb00366.x

Giddelo, C. S., Arndt, A. D., & Volckaert, F. A. M. (2002). Impact of rifting and hydrography on the genetic structure of *Clarias gariepinus* in eastern Africa. *Journal of Fish Biology*, *60*, 1252–1266. https://doi.org/10.1111/j.1095-8649.2002.tb01718.x

Goldstein, D. B., Linares, A. R., Cavalli-Sforza, L. L., & Feldman, M. W. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics*, *139*, 463–471.

Goodier, S. A. M., Cotterill, F. P. D., O'Ryan, C., Skelton, P. H., & de Wit, M. J. (2011). Cryptic diversity of African tigerfish (genus *Hydrocynus*) reveals palaeogeographic signatures of linked neogene geotectonic events. *PLoS ONE*, *6*, e28775. https://doi.org/10.1371/journal.pone.0028775

Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.). Retrieved from http://www.unil.ch/izea/softwares/fstat.html

Goudie, A. S. (2005). The drainage of Africa since the Cretaceous. *Geomorphology*, *67*, 437–456. https://doi.org/10.1016/j.geomorph.2004.11.008

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, *59*, 307–321. https://doi.org/10.1093/sysbio/syq010

Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, *2*, 618–620. https://doi.org/10.1046/j.1471-8286.2002.00305.x

Hewitt, G. M. (2004). The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology*, *2004*, 4.

Irwin, D. M., Kocher, T. D., & Wilson, A. C. (1991). Evolution of the cytochrome b gene of Mammals. *Journal of Molecular Evolution*, *32*, 128–144. https://doi.org/10.1007/BF02515385

Joyce, D. A., Lunt, D. H., Bills, R., Turner, G. F., Katongo, C., Duftner, N., … Seehausen, O. (2005). An extant cichlid fish radiation emerged

in an extinct Pleistocene Lake. *Nature*, *435*, 90–95. https://doi.org/10.1038/nature03489

Katongo, C., Koblmüller, S., Duftner, N., Mumba, L., & Sturmbauer, C. (2007). Evolutionary history and biogeographic affinities of the serranochromine cichlids in Zambian rivers. *Molecular Phylogenetics and Evolution*, *45*, 326–338. https://doi.org/10.1016/j.ympev.2007.02.011

Klett, V., & Meyer, A. (2002). What, if anything, is a Tilapia? – Mitochondrial ND2 phylogeny of Tilapiines and the evolution of parental care systems in the African cichlid fishes. *Molecular Biology and Evolution*, *19*, 865–883. https://doi.org/10.1093/oxfordjournals.molbev.a004144

Kuhner, M. K. (2009). Coalescent genealogy samples: Window into population history. *Trends in Ecology and Evolution*, *24*, 86–93. https://doi.org/10.1016/j.tree.2008.09.007

Larmuseau, M. H. D., Raeymaekers, J. A. M., Hellemans, B., Van Houdt, J. K. J., & Volckaert, F. A. M. (2010). Mito-nuclear discordance in the degree of population differentiation in a marine goby. *Heredity*, *105*, 532–542. https://doi.org/10.1038/hdy.2010.9

Lefort, V., Longueville, J. E., & Gascuel, O. (2017). SMS: Smart model selection in PhyML. *Molecular Biology and Evolution*, *34*, 2422–2424. https://doi.org/10.1093/molbev/msx149

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

Malinsky, M., & Salzburger, W. (2016). Environmental context for understanding the iconic adaptive radiation of cichlid fishes in Lake Malawi. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(42), 11654–11656. https://doi.org/10.1073/pnas.1614272113

Malinsky, M., Svardal, H., Tyers, A. M., Misika, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology and Evolution*, *2*, 1940–1955. https://doi.org/10.1038/s41559-018-0717-x

Manni, F., Guerard, E., & Heyer, E. (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm. *Human Biology*, *76*, 173–190. https://doi.org/10.1353/hub.2004.0034

Maslin, M. A., Brierley, C. M., Milner, A. M., Shultz, S., Trauth, M. H., & Wilson, K. E. (2014). East African climate pulses and early human evolution. *Quaternary Science Reviews*, *101*, 1–17. https://doi.org/10.1016/j.quascirev.2014.06.012

McCarthy, T. S. (1983). Evidence for the former existence of a major, Southerly flowing river in Griqualand West. *Transactions of the Geological Society of South Africa*, *86*, 37–49.

McKelvey, K. S., & Schwartz, M. K. (2005). DROPOUT: A program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. *Molecular Ecology Notes*, *5*, 716–718. https://doi.org/10.1111/j.1471-8286.2005.01038.x

Miller, M. P. (2005). Alleles In Space (AIS): Computer software for the joint analysis of individual spatial and genetic information. *Journal of Heredity*, *96*, 722–724. https://doi.org/10.1093/jhered/esi119

Moore, A. E., & Larkin, P. A. (2001). Drainage evolution in south-central Africa since the breakup of Gondwana. *South African Journal of Geology*, *104*, 47–68. https://doi.org/10.2113/104.1.47

Munyikwa, K., Van Den Haute, P., Vandenberghe, D., & De Corte, F. (2000). The age and palaeoenvironmental significance of the Kalahari Sands in western Zimbabwe: A thermoluminescence reconnaissance study. *Journal of African Earth Sciences*, *30*, 941–956. https://doi.org/10.1016/s0899-5362(00)00062-2

Peck, E., & Thieme, M. (2005). Sudanic-Congo (Oubangui). In M. L. Thieme, R. Abell, N. Burgess, World Wildlife Fund, B. Lehner, E. Dinerstein, D. Olson, & P. Skelton (Eds.), *Freshwater ecoregions of Africa and Madagascar: A conservation assessment* (pp. 234–236). Island Press, Washington.

Pinton, A., Agnèse, J.-F., Paugy, D., & Otero, O. (2013). A large-scale phylogeny of *Synodontis* (Mochokidae, Siluriformes) reveals the influence of geological events on continental diversity during the Cenozoic. *Molecular Phylogenetics and Evolution*, *66*, 1027–1040. https://doi.org/10.1016/j.ympev.2012.12.009

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

QGIS Development Team (2009). *QGIS Geographic Information System*. Open Source Geospatial Foundation. http://qgis.osgeo.org

Radhakrishnan, K. V., Lan, Z. J., Zhao, J., Qing, N., & Huang, X. L. (2011). Invasion of the African sharp-tooth catfish *Clarias gariepinus* (Burchell, 1822) in South China. *Biological Invasions*, *13*, 1723–1727. https://doi.org/10.1007/s10530-011-0004-0

Rambaut, A. (2014). *FigTree*. Tree Figure Drawing Tool, Version, 1.4.2. http://tree.bio.ed.ak.uk

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, *86*, 248–249. https://doi.org/10.1093/oxfordjournals.jhered.a111573

Ree, R. H., & Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, *57*, 4–453.

Rognon, X., Teugels, G. G., Guyomard, R., Galbusera, P., Andriamanga, M., Volckaert, F., & Agnèse, L. F. (1998). Morphometric and allozyme variation in the African catfishes *Clarias gariepinus and C. anguillaris*. *Journal of Fish Biology*, *53*, 192–207. https://doi.org/10.1111/j.1095-8649.1998.tb00120.x

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (2011). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Roodt-Wilding, R., Swart, B. L., & Impson, N. D. (2010). Genetically distinct Dutch-domesticated *Clarias gariepinus* used in aquaculture in southern Africa. *Southern African Journal of Aquatic Sciences*., *35*, 241–250. https://doi.org/10.2989/16085914.2010.538507

Schmidt, H. A., Strimmer, K., Vingron, M., & von Haeseler, A. (2002). TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, *18*, 502–504. https://doi.org/10.1093/bioinformatics/18.3.502

Schultheiss, R., Van Boxlaer, B., Riedel, F., von Rintelen, T., & Albrecht, C. (2014). Disjunct distributions of freshwater snails testify to a central role of the Congo system in shaping biogeographic patterns in Africa. *BMC Evolutionary Biology*, *14*, 1471–2148. https://doi.org/10.1186/1471-2148-14-42

Scott, L. (2005). Bangweulu-Mweru. In M. L. Thieme, R. Abell, M. L. J. Stiassny, P. Skelton, B. Lehner, G. G. Teugels, & D. Olson (Eds.), *Freshwater ecoregions of Africa and Madagascar: a conservation assessment* (pp. 185–186). Washington: Island Press.

Skelton, P. H. (1994). Diversity and distribution of freshwater fishes in East and Southern Africa. *Annales-Musée Royal De L'afrique Centrale*, *Sciences Zoologiques*, *275*, 95–131.

Skelton, P. H. (2001). *A complete guide to the freshwater fishes of Southern Africa*. Cape Town, RSA: Struik Publishers.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, *139*, 457–462.

Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, *35*, 627–632. https://doi.org/10.2307/2413122

Snoeks, J., De Vos, L., & Thys van den Audenaerde, T. (1997). The ichthyogeography of Lake Kivu. *South African Journal of Science*, *93*, 579–584.

Snoeks, J., & Getahun, A. (2013). African fresh and brackish water fish biodiversity and their distribution: more unknowns than knowns. In J. Snoeks, & A. Getahun (Eds.), *Proceedings of the 4th international conference on African fish and fisheries* (pp. 69–76). Tervuren, Belgium: Royal Museum for Central Africa.

Stankiewicz, J., & de Wit, M. J. (2006). A proposed drainage evolution model for Central Africa - Did the Congo flow east? *Journal of African Earth Sciences*, *44*, 75–84. https://doi.org/10.1016/j.jafrearsci.2005.11.008

Sungani, H., Ngatunga, B. P., Koblmüller, S., Mäkinen, T., Skelton, P. H., & Genner, M. J. (2017). Multiple colonisations of the Lake Malawi catchment by the genus *Opsaridium* (Teleostei: Cyprinidae). *Molecular Phylogenetics and Evolution*, *107*, 256–265. https://doi.org/10.1016/j.ympev.2016.09.027

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585–595.

Tedesco, P. A., Oberdorff, T., Lasso, C. A., Zapata, M., & Hugueny, B. (2005). Evidence of history in explaining diversity patterns in tropical riverine fish. *Journal of Biogeography*, *32*, 1899–1907. https://doi.org/10.1111/j.1365-2699.2005.01345.x

Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram Estimation. *Genetics*, *132*, 619–633.

Thieme, M. L., Abell, R., Stiassny, M. L. J., Skelton, P., Lehner, B., Teugels, G. G., … Olson, D. (2005). *Freshwater ecoregions of Africa and Madagascar: A conservation assessment*. Washington DC: Island Press.

Trauth, M. H., Larrasoaña, J. C., & Mudelsee, M. (2009). Trends, rhythms and events in Plio-Pleistocene African climate. *Quaternary Science Reviews*, *28*, 399–411. https://doi.org/10.1016/j.quascirev.2008.11.003

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*, 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Van Steenberge, M., Gajdzik, L., Chilala, A., Snoeks, J., & Vreven, E. (2014). *Labeo rosae* (Cypriniformes: Cyprinidae) in the Congo basin: A relict distribution or a historical introduction? *Journal of Fish Biology*, *5*, 1733–1738. https://doi.org/10.1111/jfb.12491

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*, 289–384.

Vanhove, M. P. M., Van Steenberge, M., Dessein, S., Volckaert, F. A. M., Snoeks, J., Huyse, T., & Pariselle, A. (2013). Biogeographical implications of Zambezian *Cichlidogyrus* species (Platyhelminthes: Monogenea: Ancyrocephalidae) parasitizing Congolian cichlids. *Zootaxa*, *3608*, 398–400. https://doi.org/10.11646/zootaxa.3608.5.8

Verheyen, E., Salzburger, W., Snoeks, J., & Meyer, A. (2003). Origin of the superflock of cichlid fishes from Lake Victoria. *East Africa. Science*, *300*, 325. https://doi.org/10.1126/science.1080699

Villesen, P. (2007). FaBox: An online toolbox for fasta sequences. *Molecular Ecology Notes*, *7*, 965–968. https://doi.org/10.1111/j.1471-8286.2007.01821.x

Vitule, J. R., Umbria, S. C., & Aranha, J. M. R. (2006). Introduction of the African catfish *Clarias gariepinus* (Burchell, 1822) into Southern Brazil. *Biological Invasions*, *8*, 677–681. https://doi.org/10.1007/s10530-005-2535-8

Watchman, A. R., & Tweddle, C. R. (2002). Relative and 'absolute' dating of land surfaces. *Earth Science Reviews*, *58*, 1–49. https://doi.org/10.1016/s0012-8252(01)00080-0

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Yu, Y., Harris, A. J., Blair, C., & He, X. J. (2015). RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography. *Molecular Phylogenetics and Evolution*, *87*, 46–49. https://doi.org/10.1016/j.ympev.2015.03.008

Yu, Y., Harris, A. J., & He, X. J. (2010). S-Diva (statistical dispersal-vicariance analysis): A tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution*, *56*, 848–850. https://doi.org/10.1016/j.ympev.2010.04.011

Yue, G. H., Kovacs, B., & Orban, L. (2003). Microsatellites from *Clarias batrachus* and their polymorphism in seven additional catfish species. *Molecular Ecology Notes*, *3*, 465–468. https://doi.org/10.1046/j.1471-8286.2003.00486.x

Zhao, K., Duan, Z. Y., Peng, Z. G., Guo, S. C., Li, J. B., He, S. P., & Shao, X. Q. (2009). The youngest split in sympatric schizothoracine fish (Cyprinidae) is shaped by ecological adaptations in a Tibetan Plateau glacier lake. *Molecular Ecology*, *18*, 3616–3628. https://doi.org/10.1111/j.1365-294x.2009.04274.x

BIOSKETCH

This study represents a collaboration between research teams that all study African freshwater biota, but that differ in expertise and/or geographical scope. The different teams are centres of expertise in fish taxonomy and biogeography (RMCA, RBINS), fish genetics and fish parasitology (KU Leuven, U Graz, U Hasselt, U Stellenbosch), and the functioning of Central (U Namur, U Lubumbashi) and Southern (U Stellenbosch) African freshwater systems. The geographical coverage of our dataset was obtained by combining data collected during four decades of fieldwork throughout the continent by the various research teams. The lead author is an evolutionary fish taxonomist who has worked as a researcher in several of the research groups.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Van Steenberge MW, Vanhove MPM, Chocha Manda A, et al. Unravelling the evolution of Africa's drainage basins through a widespread freshwater fish, the African sharptooth catfish *Clarias gariepinus*. *J Biogeogr.* 2020;47:1739–1754. https://doi.org/10.1111/jbi.13858