Received: 28 October 2021 | Revised: 10 April 2022 | Accepted: 4 May 2022

DOI: 10.1111/mec.16495

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Riverscape genomics of cichlid fishes in the lower Congo: Uncovering mechanisms of diversification in an extreme hydrological regime

Naoko P. Kurata[1,2] | Michael J. Hickerson[1,3,4] | Sandra L. Hoffberg[5] | Ned Gardiner[6] | Melanie L. J. Stiassny[2,7] | S. Elizabeth Alter[2,8]

[1]The Graduate Center of the City University of New York, New York, New York, USA

[2]Department of Ichthyology, American Museum of Natural History, New York, New York, USA

[3]The City College of New York, New York, New York, USA

[4]Division of Invertebrate Zoology, American Museum of Natural History, New York, New York, USA

[5]Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, New York, USA

[6]Department of Geography, University of Georgia, Athens, Georgia, USA

[7]The Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York, USA

[8]Department of Biology and Chemistry, California State University Monterey Bay, Seaside, California, USA

Correspondence
Naoko P. Kurata, The Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016, USA.
Email: kurata.naoko.jp@gmail.com

Funding information
National Science Foundation/Division of Environmental Biology Award, Grant/Award Number: 1655694; The CUNY Graduate Center Mina Rees Dissertation Fellowship 2021-2022; THE CUNY Graduate Center Dissertation Year Award 2020-2021; Axelrod Research Curatorship, American Museum of Natural History

Handling Editor: Katie Lotterhos

Abstract

Freshwater fishes are notably diverse, given that freshwater habitat represents a tiny fraction of the earth's surface, but the mechanisms generating this diversity remain poorly understood. Rivers provide excellent models to understand how freshwater diversity is generated and maintained across heterogeneous habitats. In particular, the lower Congo River (LCR) consists of a dynamic hydroscape exhibiting extraordinary aquatic biodiversity, endemicity, morphological and ecological specialization. Previous studies have suggested that the numerous high-energy rapids throughout the LCR form physical barriers to gene flow, thus facilitating diversification and speciation, generating ichthyofaunal diversity. However, this hypothesis has not been fully explored using genome-wide SNPs for fish species distributed across the LCR. Here, we examined four lamprologine cichlids endemic to the LCR that are distributed along the river without range overlap. Using genome-wide SNP data, we tested the hypotheses that high-energy rapids serve as physical barriers to gene flow that generate genetic divergence at interspecific and intraspecific levels, and that gene flow occurs primarily in a downstream direction. Our results are consistent with the prediction that powerful rapids sometimes act as a barrier to gene flow but also suggest that, at certain temporal and spatial scales, they may provide multidirectional dispersal opportunities for riverine rheophilic cichlid fishes. These results highlight the complexity of diversification processes in rivers and the importance of assessing such processes across different riverscapes.

KEYWORDS
Allopatric Speciation, Diversification, Gene Flow, 2RAD/3RAD, Riverscape Genomics

1 | INTRODUCTION

Despite inhabiting less than 1% of the earth's aquatic habitat, freshwater fishes are notably diverse, comprising >40% of teleost fish species (Cohen, 1970; Dawson, 2012; Horn, 1972; Lévêque et al., 2008; Miller, 2021). However, the evolutionary mechanisms underlying this "freshwater fish paradox" (Tedesco et al., 2017) remain unclear. Determining how biotic and abiotic features shape spatial patterns of genetic variation is key to understanding speciation processes across freshwater habitats. Rivers provide excellent models for such investigations because they contain heterogeneous habitats interconnected via linear corridors. Null models of riverscape genetics can be useful for testing whether genetic patterns are shaped by strictly neutral processes based on river architecture. River model networks indicate that genetic diversity accumulates unidirectionally, suggesting that populations in headwaters would display lower genetic diversity than downstream populations (Morrissey & de Kerckhove, 2009; Thomaz et al., 2016). Models also suggest that the length and number of tributaries are positively correlated with the degree of genetic differentiation (i.e., F_{ST}) (Thomaz et al., 2016). Additionally, isolation-by-distance (IBD) is an important null model predicting genetic similarity decays with increasing geographic distance (Wright, 1943). Such null models can be used to investigate mechanisms of diversification in rivers.

Deviations from the null models may be detected when there are biotic or abiotic forces disrupting patterns of unidirectional genetic accumulation, differentiation, and/or IBD. Various abiotic barriers that have been reported as drivers of genetic divergence in riverine fishes include high-energy rapids (Alter et al., 2015, 2017; Markert et al., 2010; Schwarzer et al., 2011), waterfalls (Kanno et al., 2011; Lujan et al., 2020), environmental gradients (Brauer et al., 2018; Tobler et al., 2008), and artificial barriers such as dams (Hansen et al., 2014; Samarasin et al., 2017; Vega-Retter et al., 2020). Such features in riverscapes often generate genetic structure in freshwater fishes (Lundberg et al., 2000; Thomaz et al., 2016). Therefore, testing deviations from null models is an important initial step for detecting abiotic and biotic drivers shaping genetic patterns across riverscapes.

The lower Congo River (LCR) is an excellent natural laboratory to understand the evolutionary processes of speciation in freshwater fishes. The LCR represents a corridor with a few large tributaries spanning approximately 500 river kilometres from the outflow of Pool Malebo to its mouth at the Atlantic Ocean (Figure 1; Figure S1 greyscale figure). Compared to the well-studied lacustrine cichlid radiations in the East African Great Lakes, diversification processes among riverine cichlids are poorly known (Alter et al., 2017; Ford et al., 2019; Schwarzer et al., 2011). A massive volume of hydrologic discharge from the vast Congo Basin estimated at 40,662 m^3/s (Alsdorf et al., 2016) amasses at the outflow of Pool Malebo and descends 280 m through a series of confined gorges, generating high-energy rapids, cataracts, eddies, and deep pools (Alter et al., 2015, 2017; Jackson et al., 2009; Markert et al., 2010; Oberg et al., 2008; Robert, 1946; Stiassny & Alter, 2021). A 2008 survey by the U.S. Geological Survey (USGS) and the American Museum of Natural History (AMNH) revealed that the LCR is characterized by deep underwater canyons and intermittently variable bedrock bathymetry, which is expected to have a substantial influence on the underwater flow regime. The maximum depth recorded in the LCR is over 200 m, making this one of the deepest rivers in the world (Jackson et al., 2009; Oberg et al., 2008). Given that this depth profile is more similar to some of the African Great Lakes than the other river systems, the LCR provides a unique opportunity to understand diversification processes.

Although it covers less than 2% of the area of the Congo Basin, the LCR exhibits extraordinary aquatic biodiversity, endemism, morphological and ecological specialization and has been described as a "species pump" (Lowenstein et al., 2011). Exclusive of the cichlid fish radiations in Lake Tanganyika, about 30% of cichlid species of the Congo River basin are found in the LCR, and about a quarter of those are regional endemics (Stiassny & Alter, 2021). Previous studies using mitochondrial DNA and microsatellite markers have suggested that some high-energy rapids in the LCR act as physical barriers to gene flow, thereby playing an important role in driving genetic divergence and shaping the diversity of the ichthyofauna (Markert et al., 2010). Furthermore, microallopatric diversification (>1.5 km) associated with rapids was detected among rheophilic cichlids in the genus *Teleogramma* using microsatellite and genome-wide SNP data (Alter et al., 2017; Markert et al., 2010). These empirical studies suggest that allopatric speciation is mediated via high-energy rapids in certain areas of the LCR (referred to here as the "rapids-driven allopatry hypothesis") (Alter et al., 2015, 2017; Markert et al., 2010; Schwarzer et al., 2011). However, a more in-depth understanding of how hydrological barriers are likely to influence the genetic divergence throughout the LCR is needed, as previous studies focused on narrow regions (Markert et al., 2010) or had substantial gaps in sampled locations (Alter et al., 2017). In addition, co-occurring species do not always respond to rapids in the same way; for example, previous population structure analysis shows that *Teleogramma depressum* and *Lamprologus tigripictilis* collected from the same geological range display dissimilar genetic breakpoints (Markert et al., 2010). However, that analysis was performed with traditional mitochondrial DNA and microsatellite markers. Therefore, an additional investigation using larger sample sizes encompassing wider regions of the LCR, and a spatially explicit genome-wide approach is needed to gain a deeper understanding of how hydrological barriers influence fish diversification processes in this hydrologically complex system.

This study investigates four rheophilic *Lamprologus* species (Ovalentaria, Cichlidae) endemic to the LCR (Figure 1). *Lamprologus* represents the cichlid genus with the greatest number of specimens available in the LCR. Members of the family Cichlidae, particularly those of the African Rift Valley lakes, are recognized as model systems for evolutionary studies in tropical freshwaters (e.g., Kocher, 2004; Salzburger, 2018; Santos & Salzburger, 2012). In Lake Tanganyika, within the Congo basin, cichlid diversity is dominated by members of the tribe Lamprologini and a number of species of *Lamprologus* are also distributed in the main channel of the Congo River (Schelly & Stiassny, 2004; Stiassny, 1997). Of the riverine *Lamprologus*, at

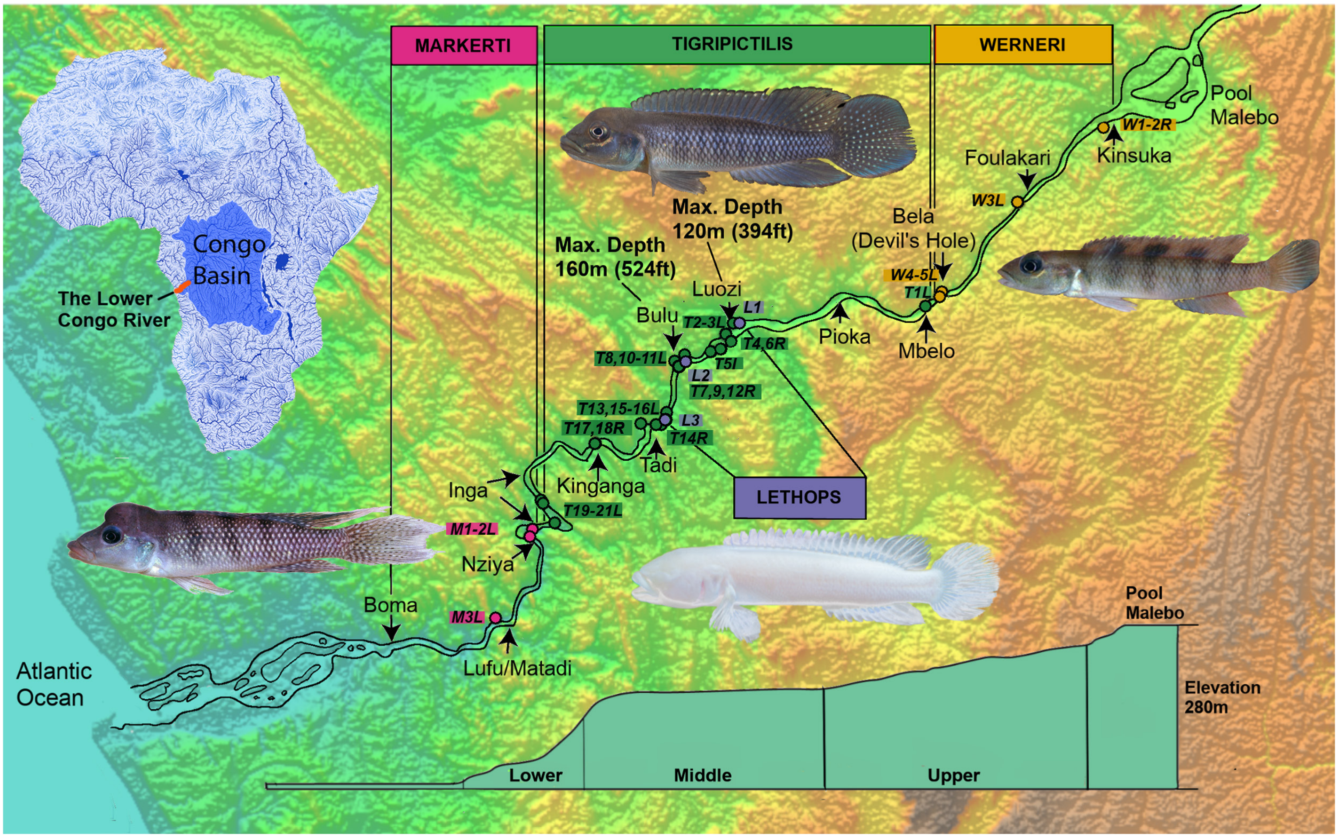

FIGURE 1 Sampling locations in the lower Congo River (LCR). Habitat distributions represented as the horizontal bars are colour-coded as *L. werneri* (W) in orange, *L. tigripictilis* (T) in green, *L. markerti* (M) in pink, and *L. lethops* (L) in purple. Whether samples are collected from the right bank (R), left bank (L), or an island (I) indicated at the end of the sample location name (e.g., W1-2R: *L. werneri* sample 1 and 2 collected from the right bank). The background image shows the elevational gradient of the LCR (Warmer tones indicate higher elevation, cooler tones indicate lower elevation). A cross-section diagram of elevational decline from east to west is shown at the bottom. *L. tigripictilis* is distributed along the longest extent of the river in comparison with the other endemic lamprologines, with a range from just below that of *L. werneri* near Mbelo to the Inga rapids (~265 river km) (Tougas & Stiassny, 2014). Below the Inga rapids, *L. tigripictilis* is replaced by *L. markerti* whose range continues downstream to Boma. *L. lethops* is another endemic species found in the Bulu-Luozi and Tadi regions of the LCR. Some areas are inaccessible for sampling due to the presence of extremely steep banks/cliffs that enclose non-navigable stretches of river due to fast-moving rapids

least four species are endemic to the LCR: *L. tigripictilis* (Schelly & Stiassny, 2004), *L. werneri* (Poll, 1959), *L. markerti* (Tougas & Stiassny, 2014), and *L. lethops* (Roberts & Stewart, 1976). These species have diversified in-situ and are distributed across numerous LCR rapids, making them an excellent study system for investigating the role of hydrology on population structure and speciation.

Lamprologus tigripictilis was first described based on morphological characteristics distinguishing it from *L. werneri*, and later, *L. markerti* was described based on genetic and morphological differences from *L. tigripictilis* (Markert et al., 2010; Schelly & Stiassny, 2004; Tougas & Stiassny, 2014). *Lamprologus lethops* is another endemic species found in the Bulu-Luozi and Tadi regions of the LCR. *Lamprologus lethops* is the only cichlid species to exhibit striking degenerative features, including an overall degeneration of the eyes and loss of pigmentation (Roberts & Stewart, 1976; Schobert et al., 2013). Available evidence suggests that *L. lethops* live at extreme depths in LCR canyons in low-light environments (Aardema et al., 2020; Stiassny & Alter, 2021). Lower Congo *Lamprologus* are small-bodied, physoclistous cichlids with limited intrinsic dispersal capabilities, particularly in deep and/or fast-flowing waters (Stiassny & Alter, 2021). Although it is currently unknown in *L. lethops*, other *Lamprologus* are pair-bonding substrate spawners, and breeding males establish and defend small territories in and around rocks where eggs are laid in caves and crevices rather than on open substrate. Females actively guard eggs and newly hatched larvae while males patrol and defend the breeding sites (A. Lamboj, personal communication, October 20, 2021). These small-bodied cichlids with limited intrinsic dispersal abilities are useful for understanding the influence of extrinsic drivers on the diversification mechanisms.

As noted by Roberts and Stewart (1976), unlike in most large African rivers where reproductive activities are strongly linked to periods of increasing water levels, the relative stability of the lower Congo results in far less seasonality, and breeding appears to occur continuously rather than being restricted to a single rainy season. While the precise habitat of *L. lethops* remains conjectural (Stiassny & Alter, 2021), *L. werneri* and *L. markerti* are most commonly found in proximity to rapids but are also found in a variety of shoreline habitats over rocky or sandy bottoms. Although also common around rapids, *L. tigripictilis* is a habitat generalist found in a wider range of shoreline habitats such as riffles, in side-channel and still water pools, over rocky, sandy, and muddy bottoms, as well as among flooded grasses (Markert et al., 2010; Roberts & Stewart, 1976). While all are primarily invertivores (Schelly & Stiassny, 2004), possible dietary partitioning between them is suggested by species-specific differences in gut length and coiling pattern (Tougas & Stiassny, 2014). Unfortunately, gut contents of preserved specimens available for study are too degraded to document possible dietary partitioning among the three species. While the generation time of Lower Congo *Lamprologus* in nature is unknown, that of *L. tigripictilis* is 14–18 months dependent on food and water temperature (O. Lucanus, personal communication, July 7, 2020). Differences in life history and the degree to which aquatic organisms are tied to their environments (e.g., water-level fluctuation) are essential components of understanding speciation processes.

Here, we generated genome-wide SNPs using reduced-representation sequencing to test the rapids-driven allopatry hypothesis (Alter et al., 2015, 2017; Markert et al., 2010; Schwarzer et al., 2011). We first inferred phylogenetic relationships across the four endemic species of *Lamprologus*. We then tested the null prediction of accumulation of downstream genetic diversity across sampling locations within each species. Next, we assessed whether high-energy rapids diminish gene flow among the species and populations of the four species with a special emphasis on the most widespread species, *L. tigripictilis*. We inferred genetic structure using spatial and nonspatial clustering approaches. Finally, we assessed interspecific and intraspecific effective migration rates and directionality of gene flow relative to river flow.

Understanding baseline genetic information is important for developing conservation management plans and understanding the potential impacts of climate change and ongoing hydroelectric dam construction in the LCR. In addition, results from our study have implications for our understanding of speciation processes and improving conservation efforts in freshwater fishes elsewhere, as global freshwater fish diversity continues to be threatened by anthropogenic impacts such as rapid climate change, hydroelectric dam construction, and overharvesting (Harrison et al., 2016).

2 | MATERIALS AND METHODS

2.1 | Sample collection

We sampled a total of 169 individuals from four riverine *Lamprologus* from the LCR: *L. tigripictilis* ($n = 112$), *L. werneri* ($n = 35$), *L. markerti* ($n = 14$), and *L. lethops* ($n = 8$) (Figure 1; Table S1). To better understand population-level processes within the LCR, we used the largest sample size of the most widely distributed species, *L. tigripictilis*. We chose available samples from the regions of Mbelo, Luozi, Bulu, Tadi, Kinganga, and Inga, spanning the known range of this species (Figure 1). The specimens of *L. lethops* were found dead or moribund; therefore, the recorded localities of our samples may not reflect their exact point of origin; however, we assume these localities are close enough to those points that we should be able to discern genetic patterns relative to the rapids.

2.2 | Generation of sequence data

We extracted genomic DNA from muscle tissue and fin clips using either the DNeasy Blood & Tissue Kit or the Gentra Puregene Tissue Kit (Qiagen), following the manufacturer's protocols. DNA concentrations were quantified using a Qubit dsDNA HS Assay (Thermo Fisher Scientific).

Restriction-site associated DNA (RAD) sequencing is a method of reduced representation DNA sequencing that samples a subset of homologous loci next to restriction enzyme cut sites across the genome of individuals (Andrews et al., 2016). The 2RAD/3RAD protocol is cost-effective and requires low amounts of input DNA (Bayona-Vásquez et al., 2019; Hoffberg et al., 2016). In addition, it ameliorates a major source of error in double-digest RADsequencing methods: PCR amplification can falsely inflate read coverage and, therefore, the accuracy of SNP calls (Andrews et al., 2016; Hoffberg et al., 2016; Tin et al., 2015). In the 2RAD/3RAD protocol, single molecules are tagged with a unique molecular index so that PCR duplicates are identified and removed in downstream analysis. We constructed libraries for 171 samples in two 96-well plates by following the Adapterama Protocol: 3RAD With Molecular IDs (Bayona-Vásquez et al., 2019). Substantial starting DNA concentrations made it unnecessary to employ a third enzyme.

Briefly, for each plate, we first normalized each DNA sample to 20 ng/μl. The restriction enzymes MspI and HindIII were chosen by performing in silico digestion simulations on the reference genome, *Neolamprologus brichardi* (NeoBri1.0), with the R package SimRAD (Lepais & Weir, 2014) (Table S2). We digested the DNA samples using MspI and HindIII-HF restriction enzymes (New England BioLabs) by incubating them for 1 h at 37°C. Immediately after the digestion, we ligated a unique combination of ClaI and HindIII adapters to each sample (Bayona-Vásquez et al., 2019) (Table S3). A portion of the ligation product was pooled and cleaned with 1.25× AMpure XP Beads (Beckman Coulter). Then, a one-cycle PCR was performed to ligate the iTru5_8N primer, which includes the 8-nt unique molecular index used to identify PCR duplicates. PCR reagents were removed with a 1.5× AMpure XP Bead clean-up. Finally, an 8 cycle PCR amplification was performed using iTru7 primers. A total of six iTru7 primers were used for each plate (Table S4). The product was cleaned with a 1.5× AMpure XP Bead and eluted in molecular-grade water. The size of pooled libraries was verified on an Agilent 2100 bioanalyser (Agilent Technologies), and then an automated size selection of 525 ± 75bp fragments was performed using a BluePippin (Sage Science). We verified and quantified the products with Agilent

2100 bioanalyser again before sequencing products on an Illumina HiSeq 2500 sequencer (Illumina, Inc.) using 150 bp paired-end reads at Genewiz, Inc.

2.3 | Demultiplexing, quality filtering, alignment, SNP calling

Raw Illumina reads were processed using the customized pipeline of IPYRAD API v.0.9.14 (Eaton & Overcast, 2020). The iTru7 barcodes were used to demultiplex reads into two plates, and then the individual samples within each plate were separated based on the internal barcode combinations of ClaI and HindIII. We set parameters to discard reads with more than 5 bases with a quality score less than the minimum quality score of 30 per read and allowed no barcode mismatches. The minimum depth required to make statistical base calls was set to 6. The filter adapter parameter was set to 3 to allow strict filtering, including poly-repeats. We allowed a maximum of 2 alleles per locus for our diploid species and filtered loci with more than 5% uncalled bases and heterozygous bases. Each sample's paired-end reads were aligned to a reference genome, *Neolamprologus brichardi*, version NeoBri1.0 (Ensembl GCA_000239395.1), an outgroup endemic to Lake Tanganyika. Two specimens of *Neolamprologus* are used in our phylogenetic analyses. Depending on the sensitivity of the analysis programs to missing data and the sample number we could include, different missingness levels were used. Minimum-individual coverages set ranged between 30% to 70% of all individuals (<30% –<70% per individual missingness), allowing for <22% missing data per locus (total missingness of <10.2%) in the variant call format (VCF) files generated for downstream analyses. Additional details are given in the sections below.

2.4 | Phylogenomic history

Maximum-likelihood phylogenetic analyses were performed on the 140,505 concatenated SNPs using IQ-TREE 2.0 (Nguyen et al., 2015). To assess variation across intraspecific tree topologies, we explored the data with three different per-individual missingness levels (<30%, <50%, and <70%). The <30% per-individual missing data set has a low amount of missing data while retaining samples from clades with fewer samples (e.g., *L. markerti*). However, since phylogenetic inferences are relatively robust with missing data (Tripp et al., 2017), we additionally explored <50% and <70% per-individual missing data sets (Table S5). We excluded sites with less than 0.01 of minor allele frequency to remove rare polymorphic and monomorphic sites. ModelFinder (Kalyaanamoorthy et al., 2017) was used to select the best-fit model via three optimality criteria; Akaike information criterion (AIC), corrected Akaike information criterion (cAIC), and Bayesian information criterion (BIC). Two lamprologine species endemic to Lake Tanganyika, *Neolamprologus modestus* and *N. christyi*, were used as outgroup species. Branch support values were estimated via 1,000 bootstrap replicates of ultrafast bootstrap approximation (UFBoot) (Hoang et al., 2018; Minh et al., 2013) and SH-like approximate likelihood ratio test (SH-aLRT) (Guindon et al., 2010).

2.5 | Spatial distribution of genetic diversity (π) and genetic structure

Genetic diversity was estimated by average pairwise differences (π; Tajima, 1983) using the POPULATIONS program implemented in STACKS v.2.41 (Catchen et al., 2013). We compared genetic diversity (π) among the sampling locations within each species (Table S6), and also calculated the median value of this summary statistic for each species to compare across species.

To explore genetic structure, we used software sNMF implemented in the R package LEA v.2.8.0 (Frichot & François, 2015; Frichot et al., 2014) and PCADAPT v.4.3.3 for only *L. tigripictilis* (Luu et al., 2017), and CONSTRUCT v.1.0.4 (Bradburd et al., 2018). To perform sNMF and CONSTRUCT analyses, 33,281 putatively unlinked loci (SNPs) were selected from the VCF file with <40% per-individual missing data: *L. tigripictilis* (n = 103), *L. werneri* (n = 35), *L. markerti* (n = 10), *L. lethops* (n = 8), by choosing one SNP within blocks of 1,000 bases using VCFTOOLS thin option (Danecek et al., 2011).

The program sNMF finds putative population clusters by calculating admixture coefficients using sparse non-negative matrix factorization algorithms, which are robust to departures from Hardy-Weinberg equilibrium (Frichot et al., 2014). We tested the best-fit value of the ancestral population (K) based on the number of sampling regions for each species: *L. tigripictilis* ranging from 1 to 6 (6 regions), *L. werneri* ranging from 1 to 4 (4 regions), *L. markerti*, and *L. lethops* ranging from 1 to 3 (3 regions). We performed 10 repetitions for each K and tested the alpha regularization parameter values 1, 10, 100, and 1000 to find the lowest cross-entropy value, which determines the best-fitted K value for each species. Additionally, principal component analyses (PCA) were performed to investigate the population-level divergence of the *L. tigripictilis* samples. We used PCADAPT implemented in R (Luu et al., 2017), which employs the Mahalanobis distance as a test statistic for detecting genetic clusters. We limited the K value to 6 corresponding to the six regions from which *L. tigripictilis* samples were collected.

A pattern of IBD (Wright, 1943) or a continuous pattern of population structure may be mistaken for a pattern of discrete genetic variation (e.g., physical barriers), particularly when sampling is not even throughout a species' range. For example, sampling throughout the LCR has been hindered by the inaccessibility of some locations (e.g., between Mbelo and Luozi). This can make it more difficult to interpret patterns and causes of population divergence accurately as the continuous pattern may appear as a discrete pattern due to the absence of samples. To avoid inaccurate interpretation of population structure due to sampling artifacts, we compare the fit of a discrete ("Barrier") model of population structure with continuous models of IBD using cross-validation and layer contribution methods implemented in the program CONSTRUCT (Bradburd et al., 2018). CONSTRUCT estimates ancestry proportions in a set of two-dimensional layers for continuous and discrete patterns of population structure. The rate at which relatedness decays with distance is estimated within each layer (Bradburd et al., 2018).

The cross-validation method compares spatial and nonspatial models where, in the case of pure IBD, a spatial model with no population structure ($K = 1$) would be strongly supported against a nonspatial model ($K > 1$). The same VCF files used for the sNMF analyses were converted to STRUCTURE format using the POPULATIONS program in STACKS (Catchen et al., 2013) and then to CONSTRUCT input format using a custom conversion script (Puckett, 2018). The geographic distances between sampling locations were calculated using the R package FOSSIL (Vavrek, 2011). The ranges of K values were set as $K = 1–6$ for *L. tigripictilis*, $K = 1–5$ for *L. werneri*, $K = 1–3$ for *L. markerti* and *L. lethops*, based on the number of sampling regions. We used 10 cross-validation replicates per K value, 100,000 MCMC iterations per replicate, and a 90% training proportion. Layer contributions were also calculated to evaluate the model selections of the cross-validation analysis by setting a threshold of 0.01. Additionally, we compared genetic distance (pairwise F_{ST}) generated in STACKS and the geographic distance matrices to examine the pattern of IBD.

2.6 | Interspecific and intraspecific effective migration and gene flow

To further investigate if rapids impede or facilitate gene flow, we visualized spatial population structure with estimated effective migration surfaces (EEMS) (Petkova et al., 2016) and inferred rates and directionality of gene flow using G-PHOCS (Gronau et al., 2011). Numerous rapids (i.e., ~66) varying in size occur across the LCR (Robert, 1946; Runge, 2007). We mapped known rapids using the results from the previous habitat mapping effort using satellite imagery: the Advanced Spaceborne Thermal Emission Reflection Radiometer (ASTER), the Landsat Enhanced Thematic Mapper Plus (ETM+), and the Shuttle Radar Topography Mission (SRTM) (Ridgeway, 2006; table 5.1). Rapids and turbulence are approximated by the presence of whitewater, which is influenced by multiple factors such as sediment particles, thalweg depth, river width, seasonal water flow, gradient, and precipitation (Cook et al., 2011; Dionne et al., 2008; Ridgeway, 2006; Torterotot et al., 2014). Then, we superimposed the known locations of rapids over the estimated effective migration surfaces to determine the role that rapids might play in the riverscape. We also used an image of a digital elevation model (DEM) to show the topographic surface of the LCR.

EEMS visualizes spatially explicit effective migration rates and deviation from exact IBD across riverscape. This method can be used to assess the effect of known physical barriers to genetic differentiation using georeferenced samples. For example, we can determine if the known locations of rapids correspond with regions of higher-than-average effective migration (i.e., promoters), lower-than-average effective migration (i.e., barriers), or simple IBD. If the "rapids-driven allopatry hypothesis" (Alter et al., 2015, 2017; Markert et al., 2010; Schwarzer et al., 2011) is supported, then regions of lower effective migration should correspond with locations of known rapids. EEMS analyses were applied to two data sets. First, we investigated the role of rapids in species-level genetic divergence between *L. werneri*, *L. tigripictilis*, and *L. markerti*. Next, an EEMS analysis was performed at the intraspecific level to understand how rapids play a role in the population-level divergence of the most widely distributed species, *L. tigripictilis*. In order to increase the sample number, we included additional samples by allowing more per-individual missingness: <45% for the interspecific level: *L. tigripictilis* ($n = 104$), *L. werneri* ($n = 35$), *L. markerti* ($n = 10$), and <55% for the intraspecific level: *L. tigripictilis* ($n = 108$), and kept unlinked 33,281 SNPs. An average genetic dissimilarity matrix was generated using bed2diffs_v1 (v1 ignores the missing genotypes) (Petkova et al., 2016). We used a deme size of 1,000 for the species-level analysis and 2,000 for the intraspecific level analysis of *L. tigripictilis* to compensate for the narrow dimensions of the river. We ran three independent chains for 1,000,000 MCMC iterations, each with a burnin of 200,000 using a random seed. The posterior probability trace of the combined three independent runs was checked to confirm convergence.

To examine the directionality of gene flow, we used the Bayesian MCMC demographic inference program G-PHOCS (Gronau et al., 2011). In a river with such a strong downstream flow (maximum water velocities of >4 m/s, Oberg et al., 2008), we hypothesized that gene flow would occur unidirectionally in a downstream direction. We analysed two data sets representing interspecific and intraspecific levels (Table S7). The interspecific level G-PHOCS analysis was performed to investigate whether introgression occurred unidirectionally or bidirectionally. We estimated interspecific migration rates between neighbouring species among *L. tigripictilis* from Mbelo, *L. tigripictilis* from Luozi, *L. werneri*, and *L. markerti*. Two populations of *L. tigripictilis* (Mbelo and Luozi) were included in this analysis since our phylogenetic inference results indicate that those two populations are highly divergent. Next, we estimated intraspecific migration rates between putative populations of *L. tigripictilis* from the four major sampling regions, Mbelo, Luozi, Bulu, and Inga. For each G-PHOCS analysis, we performed 1,000,000 MCMC iterations sampling every 100, discarding the first 10% burnin iterations, which are also used for the automatic search for finetune parameters. We set prior model parameters, for the species-level run, as a Gamma distribution with $\alpha = 1.0$ and $\beta = 1{,}000$ for τ and θ, $\alpha = 0.002$ and $\beta = 0.00001$ for migration rates. For the population-level run, the parameters were set as $\alpha = 1.0$ and $\beta = 10{,}000$ for τ and θ, $\alpha = 0.001$ and $\beta = 0.00001$ for migration rates. TRACER v1.6 (Rambaut et al., 2014) was used to view the MCMC traces to ensure convergence.

3 | RESULTS

3.1 | Sequencing and bioinformatics

We generated reduced representation sequencing data for 171 samples using the 2RAD method (Bayona-Vásquez et al., 2019; Hoffberg et al., 2016), which resulted in a total of

313,024,030 raw reads and 250,119,866 reads after initial quality filtering, including removal of PCR duplicates. The number of reads per individual ranged from 38,968 to 5,399,451 with a mean ± SD = 1.8 M ± 1.2 M (Table S8).

3.2 | Phylogenomic history

Species-level phylogenetic relationships of the riverine *Lamprologus* were inferred using maximum-likelihood methods, IQ-TREE (Nguyen et al., 2015). A comparison of the various missing data levels revealed that the allowed amount of missing data per individual was positively correlated with parsimony-informative sites and negatively correlated with constant sites (Table S5). The substitution model TVMe+R3 was selected for <30% and <50% per-individual missing data by all optimality criteria, BIC, AIC, and cAIC via IQ-TREE (Nguyen et al., 2015). However, the higher per-individual missingness <70% resulted in the incongruence model selection: TVMe+R3 model by BIC, but SYM+R3 model by AIC, and cAIC (Table S5).

All phylogenetic inferences with three levels of missing data <30%, <50%, <70% per individual (overall missingness is all less than 10.2%, 140,505 concatenated SNPs) strongly supported an identical overall topology among the four endemic LCR species (Figure 2: tree with <30% missing data per individual; Figures S2–S4). There is a large split between the *L. lethops* clade and the remaining species, *L. werneri*, *L. tigripictilis*, and *L. markerti* (UFBoot =100, SH-aLRT = 100). *Lamprologus werneri* diverges earlier than two sister taxa, *L. tigripictilis*, and *L. markerti*, as expected given the geographic range of *L. werneri* upstream of both *L. tigripictilis* and *L. markerti*.

Maximum-likelihood inferences also suggest population-level divergence in all species, including the widely distributed *L. tigripictilis*. Several intraspecific clades corresponding to geographic regions remained consistent across all different levels of missingness (Figure 2; Figures S2–S4). For example, all trees with different per-individual missingness levels indicate that all *L. tigripictilis* individuals from Mbelo are isolated from all other downstream samples (UFBoot = 100, SH-aLRT = 100). Clustering of most of the Inga samples is also observed in the <30 and 50% per-individual missingness level trees (UFBoot = 100, SH-aLRT = 100); all Inga samples cluster in the <70% per-individual missingness level tree (UFBoot = 100, SH-aLRT = 95). Several other population-level subclades also reflect the proximity of locations corresponding to the geographic regions of Luozi, Bulu, Tadi, and Kinganga. For example, some individuals from two adjacent sampling localities within the region of Luozi cluster together regardless of per-individual missingness level (e.g., TIG2L and TIG3L). However, we also observed clustering of samples between adjacent regions (e.g., a few samples of TIG6R at Luozi and TIG7R at Bulu).

Within *L. werneri*, the individuals from the uppermost extent of their range at Kinsuka (right bank) diverge first, followed by individuals from Foulakari (left bank, 40.2 river km downstream of Kinsuka) and the two downstream populations from Bela (Figure 2). The upper Bela population (42.2 river km downstream of Foulakari) is monophyletic (UFBoot = 100, SH-aLRT = 100) and nested within lineages from lower Bela. This overall topology is consistent among all missingness levels (Figure 2; Figures S2–S4).

Within *L. markerti*, genetic divergence between samples from Nziya and Lufu (39.1 river km downstream of Nziya sampling point) is detected; however, there is no divergence between upper and lower Nziya which are approximately 2.1 km apart. This pattern of the tree topology remains throughout all missingness levels. *Lamprologus lethops* is monophyletic, and tree topologies are consistent regardless of the level of missingness (Figure 2; Figures S2–S4).

3.3 | Spatial distribution of genetic diversity and genetic structure

3.3.1 | Interspecific and intraspecific genetic diversity

To test the null hypothesis of downstream accumulation of genetic diversity, we calculated genetic diversity (π) for each sampling location within each species (Figure S5; Table S6). The results show no accumulation of genetic diversity for *L. werneri*, *L. tigripictilis*, and *L. markerti*. While *L. lethops* is an exception, the sample size for this species was relatively small (n = 8). Species-level genetic diversity (π) across all four species is similar to the median genetic diversity of cichlid fishes, which is in itself considered low for vertebrates (Svardal et al., 2021) (Figure S6).

3.3.2 | Genetic structure and admixture

To investigate the role of high-energy rapids in population divergence and isolation, we measured population structure and admixture within species using three methods: sNMF (Frichot & François, 2015; Frichot et al., 2014), PCADAPT (Luu et al., 2017), and CONSTRUCT (Bradburd et al., 2018). Each of these methods is approximately in agreement with the K values obtained from sNMF that delineated groups separated by rapids: $K = 2$ or 3 for *L. tigripictilis* (best-fitted K is 3, but there is only a clear separation between Mbelo and the rest of the samples, Figure 3, $K = 3$), $K = 2$ for *L. werneri*, and no structure for *L. markerti* and *L. lethops*.

The estimations of the admixture coefficients using sNMF suggest that the best-fitted K determined by the minimum-cross entropy method for *L. tigripictilis* is 3 (Figure 3; Figure S7). These genetic clusters correspond to Mbelo, and the highly admixed regions, including Luozi, Bulu, Tadi, Kinganga, Inga, are composed of two layers of ancestry proportions (Figure 3). Interestingly, however, fine-population structure corresponding to geographic regions is detectable at the higher K values (Figure 3). The genetic isolation of the Inga samples becomes clear at $K = 5$. The PCA analyses using PCADAPT (Luu et al., 2017) mostly support the clustering pattern suggested by the sNMF analyses (Figure 4).

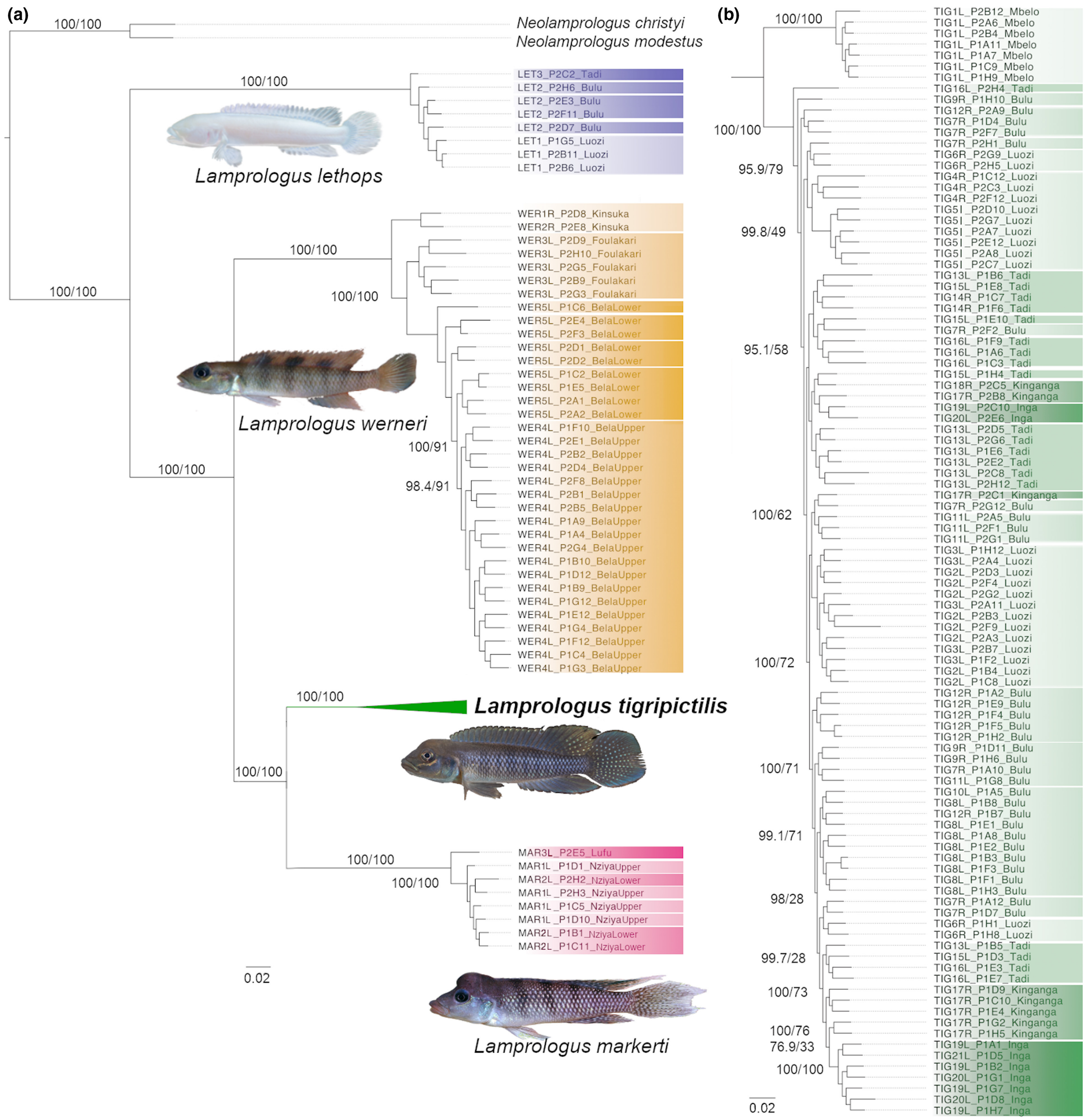

FIGURE 2 Maximum-likelihood phylogenetic inference from IQ-TREE under TVMe+R3 and default search values using 140,505 SNPs (<30% missing data per individual). (a) Species-level tree (with *L. tigripictilis* clade collapsed). (b) Expanded view of the *L. tigripictilis* clade. Light to dark colour gradients of each species corresponds to upstream to downstream sample locations. Branch support values were estimated via 1,000 bootstrap replicates of ultrafast bootstrap approximation (UFBoot) and SH-like approximate likelihood ratio test (SH-aLRT) (see detailed branch support values in Figure S2)

The cross-validation analysis of CONSTRUCT suggests the best-fitted value *K* is around 4; however, layer contributions beyond the second layers are not significant given that they are below the threshold of 0.01. Thus, it is suggested that the best *K* is 2 (Figures S8–S10; Table S9). In addition, the predictive accuracy scores of the cross-validation analysis suggest that there is no significant difference between spatial and nonspatial models for *L. tigripictilis*; therefore, IBD is not supported as the main diversification mechanism (Figure S8). This result was also supported by a simple comparison of genetic (pairwise F_{ST}) and geographic distances ($R^2 = 0.253$) (Figure S11).

The sNMF result of *L. werneri* suggests the best *K* value is 2 (Figure 3; Figure S7). The genetic break between Foulakari and upper Bela (42.2 river km downstream of Foulakari) coincides with a cross-channel rapid; however, there is no apparent genetic divergence

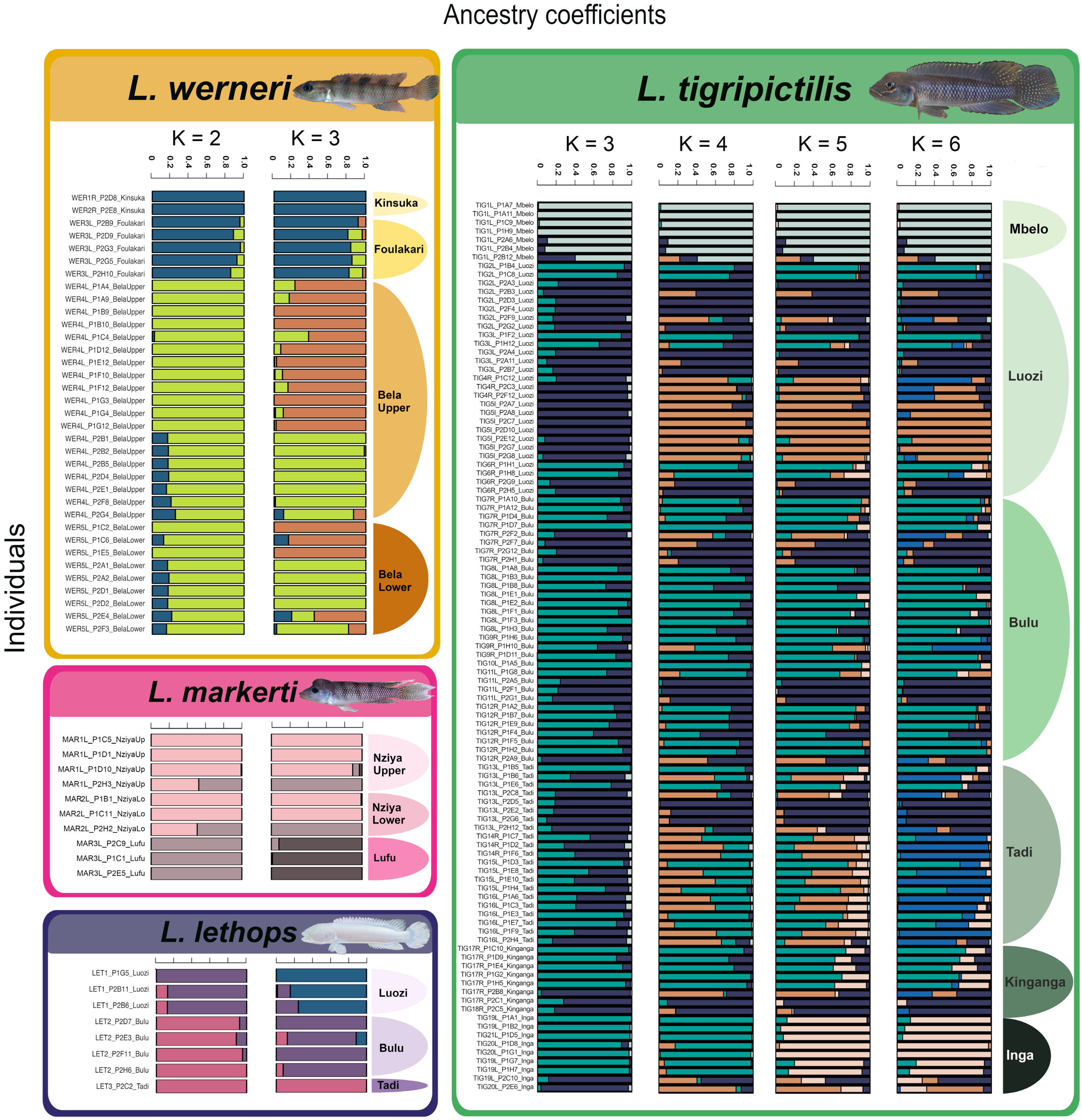

FIGURE 3 Population structure inferred from sNMF analyses of four endemic species. Each bar represents an individual sample. Individuals within each plot are arranged from upstream (top) to downstream (bottom). Whether samples are collected from the right bank (R), left bank (L), or an island (I) indicated at the end of the sample name (e.g., TIG1L: *L. tigripictilis* sample 1 collected from the left bank). *L. tigripictilis*: $K = 3$ (best-fitted value of K) indicates that Mbelo is isolated while the rest of the regions show mixed ancestry; however, greater values of K reveal finer-scale structure (e.g., substructure within the Luozi region that correspond to samples from a rocky island and the right bank near the island ($K > 4$), as well as substructure within the Inga region ($K > 5$). *L. werneri*: $K = 2$ (best-fitted K value) shows the genetic divergence between Foulakari and Bela. $K = 3$ reveals additional substructure within upper Bela and lower Bela. *L. markerti*: best-fitted K is $K = 1$; however, $K = 2$ (left plot) and $K = 3$ (right plot) indicate the divergence between Nziya and Lufu, which might be influenced by a series of large rapids between the regions (Figure 5a). *L. lethops*: best-fitted value is $K = 1$; however, $K = 2$ (left plot) and $K = 3$ (right plot) suggest geographic substructure between Luozi, Bulu and Tadi. These results may support that *L. lethops* samples are collected from the respective regions even though the exact point of origin within the region is unknown

despite the presence of the rapids between Kinsuka and Foulakari (40.2 river km downstream of Kinsuka). The cluster value $K = 2$ was also suggested by both spatial and nonspatial models in the cross-validation analysis and layer contributions in CONSTRUCT (Figures S8–S10). IBD was not detected by the comparison of genetic and geographic distances ($R^2 = 0.083$) (Figure S11).

The sNMF cross-entropy values for *L. markerti* suggest the best K value is 1 (Figure 3; Figure S7). The same K value is also suggested by the CONSTRUCT cross-validation and layer contribution analyses (Figures S8–S10). While $K = 2$ was the best K suggested by the nonspatial models, the spatial models outperformed nonspatial models suggesting no genetic structure, and the patterns may reflect IBD. The pattern of IBD was also seen in the comparison of genetic and geographic distances ($R^2 = 0.945$) (Figure S11). Within the distribution of *L. markerti*, there are a series of large rapids between Nziya and Lufu (Figure 5a). Although $K = 1$ was best supported, these rapids may have some influence on the genetic divergence between Nziya and Lufu (Figure 3: *L. markerti* $K = 3$).

We predicted genetic divergence in samples of *L. lethops* between the Luozi, Bulu, and Tadi due to the intensive upwelling of the Bulu reach (Jackson et al., 2009), which could potentially act as a physical barrier. However, $K = 1$ is best supported by the sNMF analysis. Also, the CONSTRUCT spatial models outperformed nonspatial models suggesting no population structure among those sampling locations following a pattern of IBD (Figure S8). The pattern of IBD was also seen in the comparison of genetic and geographic distances ($R^2 = 0.715$) (Figure S11). Although no population structure is best supported, the sNMF population structure $K > 1$ indicates a signal of genetic divergence corresponding to the sampling regions (Figure 3). Regardless, the wider 95% confidence intervals of the CONSTRUCT spatial cross-validation results for both *L. markerti* and *L. lethops* indicate that an increased sample number may help in generating more confidence in these results (Figure S9).

3.4 | Interspecific and intraspecific effective migration and gene flow

We performed interspecific and intraspecific level EEMS analyses to evaluate the correlation of rapids with effective migration rates. In the interspecific level analysis, we detected lower-than-average effective migration in the contact zones between the distributions of *L. werneri* and *L. tigripictilis* (Figure 5b) and between *L. tigripictilis* and *L. markerti* (Figure 5c), suggesting that the rapids/turbulence may be acting as major barriers at these locations. We additionally evaluated the direction of hybridization between species via the interspecific-level demographic inferences using G-PHOCS (Gronau et al., 2011). While we hypothesized unidirectional gene flow from the upstream to the downstream species, the results indicate bidirectional gene flow and a low level of hybridization between the neighbouring species (Figure 6a, Table S10). As expected based on genetic proximity, there is a higher gene flow between *L. tigripictilis* from Mbelo and Luozi than that between *L. werneri* and *L. tigripictilis* from Mbelo and between *L. tigripictilis* from Luozi and *L. markerti*.

In the population-level EEMS analysis of *L. tigripictilis*, we observed a mosaic of processes across the species range (Figure 5d).

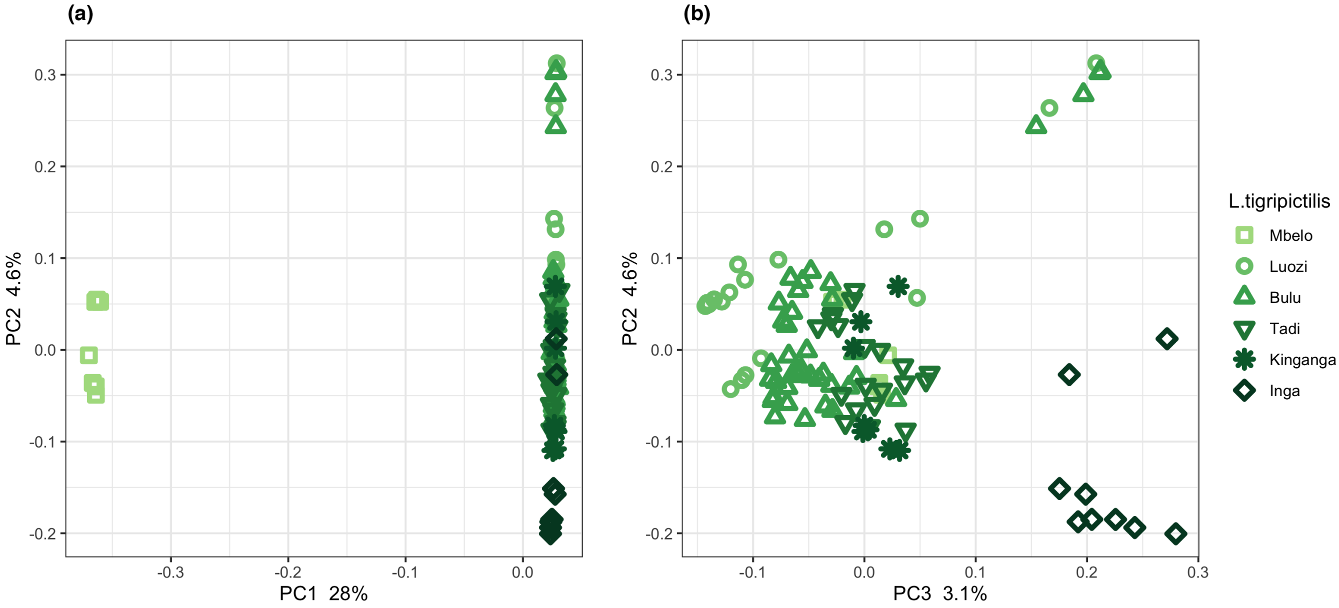

FIGURE 4 Population-level principal component analyses of *L. tigripictilis*. (a) PC1 (28%) versus PC2 (4.6%): The isolation of Mbelo corresponds to 28% of the variance explained by PC1, (b) PC2 (4.6%) versus PC3 (3.1%): The separation of the Inga samples is observed along the PC3 axis, in which 3.1% of the variance is explained. Overlapping clusters are observed for the sampling regions of Luozi, Bulu, Tadi, and Kinganga, with a few exceptions of outgroups. Symbols represents sampling regions of origin within of *L. tigripictilis*' distribution: □ = Mbelo, ◯ = Luozi, △ = Bulu, ▽ = Tadi, * = Kinganga, and ◇ = Inga. Colour gradation from lighter to darker green indicates the direction of the river flow from upstream to downstream

We detected lower-than-average effective migration surfaces in the river stretch between Mbelo and Luozi, near the bend between Luozi and Bulu, and in the Inga rapids. The results correspond to the population structure generated by sNMF (Figure 3: *L. tigripictilis*). A section of higher effective migration is observed between Bulu and Kinganga, which suggests high connectivity within this region. This result is supported by the high degree of admixture detected in the sNMF analysis (Figure 3: *L. tigripictilis* $K = 3$). The cross-channel rapids (red spirals) and partial-channel rapids (blue spirals) overlap with the lower migration surface (i.e., orange shades); however, they are also present in the region of higher-than-average effective migration (i.e., blue shades) (Figure 5d). The river reach around Luozi has no surface rapids and is largely influenced by IBD.

In addition, we estimated intraspecific-level migration rates across four populations of *L. tigripictilis* from Mbelo, Luozi, Bulu, and Inga using G-PHOCS. The analyses suggest bidirectional gene flow upstream and downstream between neighbouring populations, which is inconsistent with our hypothesis of unidirectional gene flow along the river (Figure 6b; Table S10).

4 | DISCUSSION

Understanding why freshwater fishes are so diverse compared to marine fishes, despite occupying a tiny fraction of available aquatic habitat on earth, remains a central question in vertebrate evolution. However, the mechanisms by which some rivers generate unusually high ichthyofaunal diversity and endemicity remain poorly known. Previous studies have inferred allopatric processes as a primary mechanism driving the separation of lineages within and between rivers (Boschman et al., 2021; Dias et al., 2013; Lujan et al., 2020; Miller, 2021); however, diversification in sympatry or parapatry has also been hypothesized for some groups (Burress et al., 2018; Sullivan et al., 2002) including another cichlid genus in the lower Congo (Alter et al., 2017). In this study, we investigated a previously suggested rapids-driven allopatry hypothesis in the LCR (Alter et al., 2015, 2017; Markert et al., 2010; Schwarzer et al., 2011) using genome-wide SNP data across and within four endemic lamprologine cichlids. Our results indicate that major rapids accompanied by significant elevational drops correspond to species boundaries, indicating that allopatric processes due to strong barriers are important in driving diversity in this system (Figures 2 and 5). However, patterns at the population-level are more complex, showing rapids correspond to both increased and decreased gene flow at different locales, and that in the absence of rapids, patterns are best explained by IBD (Figures 5–6; Figure S8). Moreover, the lack of higher genetic diversity at downstream sites (Figure S5), as well as evidence of multidirectional gene flow (Figures 3 and 6), highlight complex evolutionary patterns even within this relatively linear system (Svardal et al., 2021).

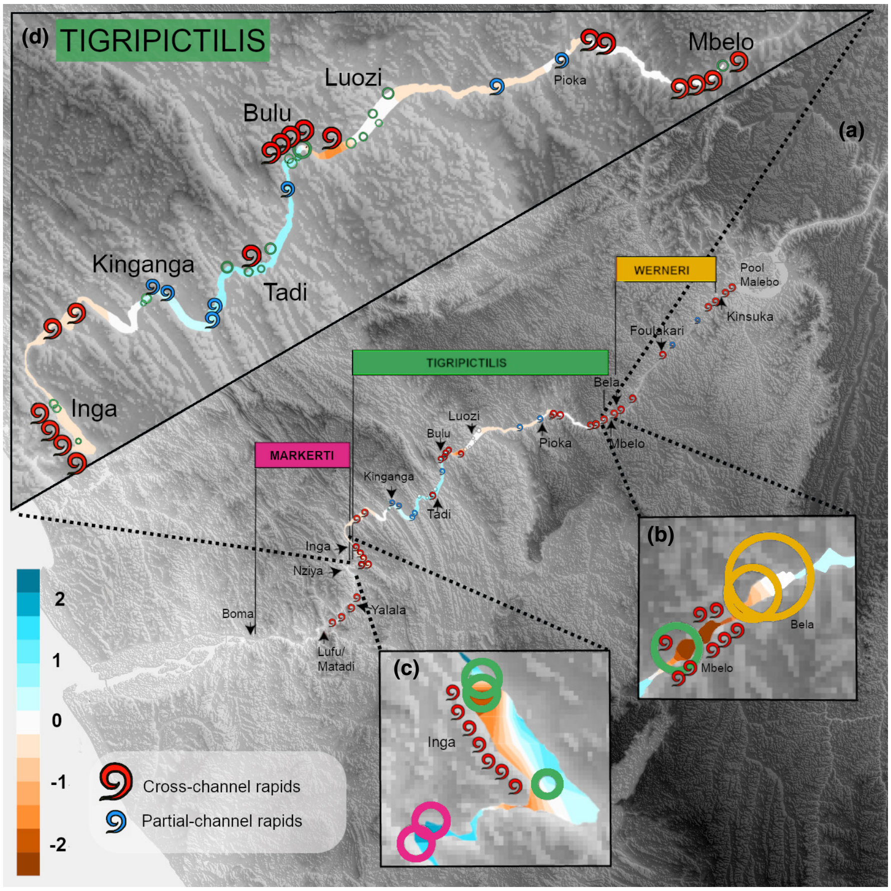

FIGURE 5 Species-level and population-level estimated effective migration surfaces (EEMS). (a) Locations of high-energy rapids in the lower Congo River: red spirals are cross-channel rapids, blue spirals are partial-channel rapids. (b) Species-level EEMS result of the contact zone between *L. werneri* (orange circles) and *L. tigripictilis* (green circle). (c) Species-level EEMS result in Inga separating *L. tigripictilis* (green circles) and *L. markerti* (pink circles). The size of the circles corresponds to the relative sample size. (d) Population-level EEMS result of *L. tigripictilis*. Shades of orange correspond to lower-than-average migration (i.e., barriers), shades of blue correspond to higher-than-average migration (i.e., promotors of gene flow), and white colour represents isolation-by-distance. The background shows the topographic surface of the lower Congo basin with darker shade corresponding to higher elevation

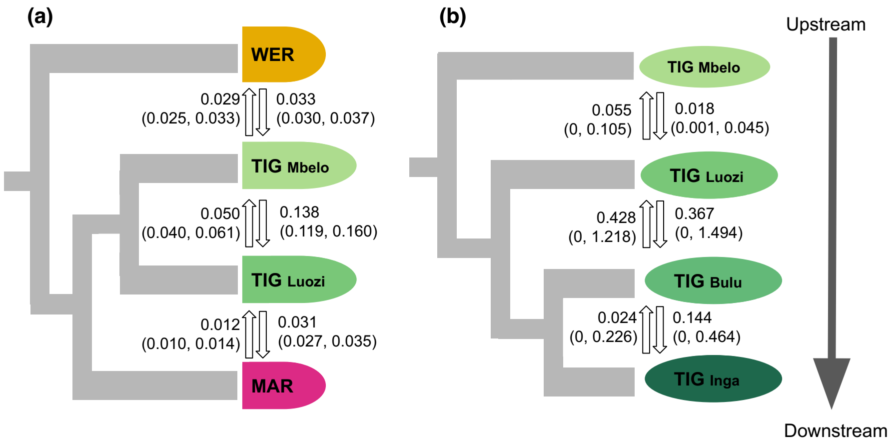

FIGURE 6 Migration rates and directionality estimated by G-PHOCS. Migration rates (mean, 95% highest posterior density intervals) are depicted as the number of migrants per generation. Species and populations are colour-coded: *L. werneri* (WER) is orange, *L. markerti* (MAR) is pink, *L. tigripictilis* (TIG) from different locations are different shades of green and colour gradation from lighter to darker green indicates the direction of the river flow from upstream to downstream. Evidence of upstream and downstream bidirectional gene flow are observed between (a) neighbouring species and (b) populations of *L. tigripictilis*. The lower bounds of the 95% HPD intervals include 0 for some pairwise *L. tigripictilis* populations. Generally, if the migration rate estimates overlap with 0, then there is no strong evidence of gene flow. However, the intervals are very wide often with high upper bounds that are consistent with a high degree of uncertainty

4.1 | Role of high-energy rapids in genetic divergence of lamprologine cichlids

Higher rates of diversification in freshwater fishes compared to marine fishes have been hypothesized to be shaped by barriers that reduce gene flow (e.g., allopatric processes), but meta-analyses of diversification mechanisms have produced mixed results (Miller, 2021). Our study suggests that major rapids correspond to species boundaries in LCR lamprologines, and therefore appear to be important drivers of speciation processes. The rapids at Inga have been suggested to form a biogeographic barrier for other cichlid taxa (Alter et al., 2015, 2017; Markert et al., 2010; Schwarzer et al., 2011). For example, the rheophilic cichlid genus *Teleogramma* is absent below the Inga rapids (Alter et al., 2015; Schwarzer et al., 2011), and the cichlid genus *Nanochromis* does not occur below Nziya (Schwarzer et al., 2011).

Across populations within the distribution of *L. tigripictilis*, the role of cross-channel and smaller rapids appear more complex. While the regions of lower-than-average migration or barriers (e.g., orange shades in Figure 5d) broadly overlap with the sections of elevational decline where rapids probably occur, our results suggest that rapids may sometimes act as barriers, but in other instances, they may promote gene flow (Figure 5d). Within the range of *L. tigripictilis*, Mbelo is the most genetically and geographically isolated group. Water velocity below Mbelo is high due to an elevational drop, and a series of rapids formed where the river repeatedly widens and narrows (Figures 1–3). Our study suggests that a large genetic break between Mbelo and Luozi is partly driven by barriers (orange shades, Figure 5d) where at least one partial-channel rapid is identified from satellite imagery. Likewise, the genetic isolation of the Inga samples is supported by our analyses (Figure 2; Figure 3 *L. tigripictilis* $K = 5$ and 6; Figure 4b), and the genetic isolation appears to be driven by the presence of cross-channel rapids (Figure 5d).

In contrast, population structure in other regions, such as Luozi, Bulu, Tadi, and Kinganga, presents a strong signal of admixture, despite the presence of multiple rapids between these locations (Figure 3). This pattern of high admixture may result from the more gentle elevational incline in this region compared with the larger elevational drops below Mbelo and around Inga (Figure 1). While small-bodied, physoclistous cichlid species have limited intrinsic dispersal capabilities in deep and/or fast-flowing waters (Stiassny & Alter, 2021), dispersal may be facilitated for such a small-bodied, eurytopic species as *L. tigripictilis* during wet seasons when river banks are flooded, particularly in regions of gentle elevational incline. Additionally, turbulent high-energy rapids may occasionally facilitate downstream and upstream bidirectional migration (Figure 6b). Interestingly, our population-level EEMS results indicate that the region of Luozi (exclusive of the middle island area) is strictly explained by IBD (Figure 5d). This stretch of river is devoid of surface rapids, is fully navigable, and is represented by numerous samples collected from both banks along the reach. This result supports the idea that the IBD may play a primary role when no hydrological barriers are present.

While allopatric speciation due to physical barriers is reported across many taxa including birds (Winker, 2021), herbivorous insects (Tishechkin, 2020), snapping shrimp (Knowlton & Weigt, 1998), and alpine plants (Boucher et al., 2016), some barriers may be temporary or ephemeral, particularly in some aquatic environments (e.g., soft barriers to coral reef fishes, Bowen et al., 2013; Rocha & Bowen, 2008; Tornabene et al., 2015). Hydrological barriers in the LCR are also expected to have some seasonal variation (e.g., wet and dry seasons) and may have been influenced by major shifts during Quaternary climatic oscillations. Therefore, exploring different time scales across hydrological regimes, rather than drawing conclusions from one temporal snapshot, is important for inferring the effects of these regimes on population structure. Our use of satellite imagery to detect rapids (Ridgeway, 2006) suggests potential applications to monitor changes in hydrological patterns of rivers and other aquatic systems across different temporal scales to better understand speciation dynamics.

4.2 | Complex evolutionary processes in the lower Congo River

Strongly directional physical forces such as river flow, marine currents, and wind can control the movement of gametes and individuals, influencing gene flow and broader patterns of diversification across a wide array of species (Bertola et al., 2020; Cowen & Sponaugle, 2009; Kling & Ackerly, 2021; Morrissey & de Kerckhove, 2009; Pringle et al., 2011). Simple models of such systems predict asymmetric gene flow and higher genetic diversity in downwind, downstream, and down-current locations compared to source populations. In rivers, we expect the accumulation of genetic diversity from the headwater to downstream locations in a one-dimensional linear stepping-stone pattern (Maruyama, 1969; Thomaz et al., 2016; Washburn et al., 2020). However, numerous factors could cause deviations from this expectation, including the presence of barriers to gene flow and corridors or other landscape complexities that promote multidirectional gene flow (Petkova et al., 2016). The LCR presents a useful test of the simplest null model of river architecture as it is a nearly linear corridor with few major tributaries, in contrast to complex dendritic river networks (Brauer et al., 2018; Levin et al., 2020; Ochoa et al., 2015). Our estimation of intraspecific genetic diversity (π) within each species indicates that there is no clear signal of downstream genetic diversity accumulation, except for in *L. lethops*. There are several potential explanations for this lack of the expected pattern. In addition to the physical barriers to migration demonstrated here, patterns of genetic diversity may also be influenced by the multidirectional gene flow we observed in this study. Our results suggest bidirectional migration between several populations of *L. tigripictilis* (Figure 6b) as well as potential cross-channel migration with gene flow between the right and left banks at Luozi, Bulu, and Kinganga (Figures 2 and 3). This multidirectional gene flow perhaps is the result of upstream corridors along banks created during the wet season, which may introduce genetic variation from downstream to upstream locations (Thomaz et al., 2016). In addition, it is also important to consider the possibility of historical episodic surges in gene flow due to rare events, for example, geological uplifts, severe droughts that change discharge dramatically, or unusual flooding. In summary, the results presented here suggest that despite its seemingly simple architecture, the LCR is better characterized as a more complex system such as in the infinite-islands model (Wright, 1931), rather than by a one-dimensional linear stepping-stone model (Maruyama, 1969). These findings add to the literature suggesting that patterns of gene flow may often be strongly influenced by complex and idiosyncratic landscape features and historical events, even in systems that appear linear (Kling & Ackerly, 2021).

5 | CONCLUSIONS

Our study extends the understanding of diversification processes influenced by high-energy rapids and provides insight into the evolutionary processes driving the remarkable diversity of fishes. While we focused on the rapids-driven allopatry hypothesis, explorations of other drivers will also be needed to fully decipher complex riverscape processes. Ongoing research includes the potential impact of geological changes and Quaternary climate fluctuation in the demographic history of the lamprologine cichlids. Although the geological history of the LCR is poorly known (Stiassny & Alter, 2021), a previous study posited that the upstream and downstream reaches of Inga were once separated by a large waterfall, which eroded over time, forming the current Inga rapids and allowing greater dispersal (Schwarzer et al., 2011). The hydrological regime, including the location and size of rapids, and the level of water discharge all probably differed in the past, particularly during glacial cycles, which might have had a large influence on gene flow. Future studies include surveying environmental variation across species' habitats to investigate ecological/sympatric speciation, which has been suggested in some cichlid clades (Barluenga et al., 2006; Elmer et al., 2014; Kautt et al., 2016; Malinsky et al., 2015; Poelstra et al., 2018). Furthermore, we suggest that it is important to compare diversification rates among riverine habitats varying different habitat complexity, for example, between the LCR and other subdrainages in the Congo Basin, in addition to comparison with other freshwater systems and marine habitats (Miller, 2021). Further examination of regional-based diversification rates may provide more nuanced insight into the complex diversification processes in the hydrologically extreme LCR revealed in the present study as well as the "freshwater fish paradox".

Our findings strongly suggest that hydrological features are important in maintaining and shaping biodiversity and therefore conserving that aspect of the LCR riverscape will be crucial. This study demonstrates that genome-wide SNP data can help improve fundamental understanding of the evolutionary processes generating riverine fish diversity, and such results may facilitate the planning of conservation management strategies, particularly as they relate to proposed dam development in this rich system.

AUTHOR CONTRIBUTIONS

Naoko P. Kurata and S. Elizabeth Alter conceived of and designed the study. Naoko P. Kurata carried out 2RAD library construction, performed all data quality/filtering analyses, data analyses and wrote the manuscript. Michael J. Hickerson provided suggestions with population genomic analyses. Sandra L. Hoffberg provided guidance with library construction using the 2RAD method. Ned Gardiner provided analytical results of highly-energy rapids. Melanie L. J. Stiassny and S. Elizabeth Alter conducted field studies including specimen collection, and Melanie L. J. Stiassny identified all specimens. All authors edited and approved of the manuscript.

ACKNOWLEDGEMENTS

We thank Dr Nathan Lujan for helpful suggestions on experimental ideas and Tobit Liyandja for providing extensive information about the lower Congo River. Linelle Abueg and Sajesh Singh provided much assistance with bioinformatics. Oliver Lucanus (Below Water, Montreal) kindly provided photographs. We appreciate Dr Bruno Melo for the helpful suggestions with the graphics of the phylogenetic tree. We are grateful to Lauren Audi and the staff at the Sackler Institute for Comparative Genomics (AMNH) for providing valuable assistance in the laboratory. We extend our appreciation to the AMNH Ichthyology team, Thomas Vigliotta, Kimberly Bernotas, Chloe Lewis, Radford Arrindell, Dr Barbara Brown† for their support with collection management, and Mohammad Faiz for managing laboratory supplies.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw sequence data are available on the NCBI Sequence Read Archive (BioProject ID PRJNA794726: see Table S1 for individual specimens accession IDs). Additional data products related to the analyses (e.g., VCF files, scripts, input and output files) are available on Zenodo (10.5281/zenodo.6416266). All fishes were collected and exported with permission of the Ministère de l'Agriculture, Secrétariat Général à l'Agriculture, Pêche et Elevage, Direction des Pêches, in the Democratic Republic of Congo. And the Ministère de la Recherché Scientifique et de l'Innovation Technique, Institut National de Recherché en Sciences Exactes et Naturelles (IRSEN) and Groupe d'études et de Recherche sur la Diversité Biologique (GERBID) in the Republic of Congo. All relevant permits are on file and available on request at the AMNH.

ORCID

Naoko P. Kurata https://orcid.org/0000-0002-4236-2646
Michael J. Hickerson https://orcid.org/0000-0002-5802-406X
Sandra L. Hoffberg https://orcid.org/0000-0002-7249-7453
Ned Gardiner https://orcid.org/0000-0002-0978-998X
Melanie L. J. Stiassny https://orcid.org/0000-0001-8220-4768
S. Elizabeth Alter https://orcid.org/0000-0003-0597-9206

REFERENCES

Aardema, M. L., Stiassny, M. L. J., & Alter, S. E. (2020). Genomic analysis of the only blind cichlid reveals extensive inactivation in eye and pigment formation genes. *Genome Biology and Evolution*, *12*(8), 1392–1406. https://doi.org/10.1093/gbe/evaa144

Alsdorf, D., Beighley, E., Laraque, A., Lee, H., Tshimanga, R., O'Loughlin, F., Mahé, G., Dinga, B., Moukandi, G., & Spencer, R. G. M. (2016). Opportunities for hydrologic research in the Congo Basin. *Reviews of Geophysics*, *54*(2), 378–409. https://doi.org/10.1002/2016RG000517

Alter, S. E., Brown, B., & Stiassny, M. L. J. (2015). Molecular phylogenetics reveals convergent evolution in lower Congo River spiny eels. *BMC Evolutionary Biology*, *15*, 224. https://doi.org/10.1186/s12862-015-0507-x

Alter, S. E., Munshi-South, J., & Stiassny, M. L. J. (2017). Genomewide SNP data reveal cryptic phylogeographic structure and microallopatric divergence in a rapids-adapted clade of cichlids from the Congo River. *Molecular Ecology*, *26*(5), 1401–1419. https://doi.org/10.1111/mec.13973

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. https://doi.org/10.1038/nrg.2015.28

Barluenga, M., Stölting, K. N., Salzburger, W., Muschick, M., & Meyer, A. (2006). Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, *439*, 719–723. https://doi.org/10.1038/nature04325

Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., Bentley, K. E., Finger, J. W., Louha, S., Troendle, N., Diaz-Jaimes, P., Mauricio, R., & Faircloth, B. C. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ*, *7*, e7724. https://doi.org/10.7717/peerj.7724

Bertola, L. D., Boehm, J. T., Putman, N. F., Xue, A. T., Robinson, J. D., Harris, S., Baldwin, C. C., Overcast, I., & Hickerson, M. J. (2020). Asymmetrical gene flow in five co-distributed syngnathids explained by ocean currents and rafting propensity. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1926), 20200657. https://doi.org/10.1098/rspb.2020.0657

Boschman, L. M., Cassemiro, F. A. S., Carraro, L., de Vries, J., Altermatt, F., Hagen, O., Hoorn, C., & Pellissier, L. (2021). South American freshwater fish diversity shaped by Andean uplift since the Late Cretaceous. *BioRxiv*. https://doi.org/10.1101/2021.05.14.444133

Boucher, F. C., Zimmermann, N. E., & Conti, E. (2016). Allopatric speciation with little niche divergence is common among alpine Primulaceae. *Journal of Biogeography*, *43*(3), 591–602. https://doi.org/10.1111/jbi.12652

Bowen, B. W., Rocha, L. A., Toonen, R. J., & Karl, S. A. (2013). The origins of tropical marine biodiversity. *Trends in Ecology & Evolution*, *28*(6), 359–366. https://doi.org/10.1016/j.tree.2013.01.018

Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, *210*(1), 33–52. https://doi.org/10.1534/genetics.118.301333

Brauer, C. J., Unmack, P. J., Smith, S., Bernatchez, L., & Beheregaray, L. B. (2018). On the roles of landscape heterogeneity and environmental variation in determining population genomic structure in a dendritic system. *Molecular Ecology*, *27*(17), 3484–3497. https://doi.org/10.1111/mec.14808

Burress, E. D., Alda, F., Duarte, A., Loureiro, M., Armbruster, J. W., & Chakrabarty, P. (2018). Phylogenomics of pike cichlids (Cichlidae: Crenicichla): the rapid ecological speciation of an incipient species flock. *Journal of Evolutionary Biology*, *31*(1), 14–30. https://doi.org/10.1111/jeb.13196

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Cohen, D. M. (1970). How many recent fishes are there? *Proceedings of the California Academy of Sciences*, *17*, 341–346.

Cook, B. D., Kennard, M. J., Real, K., Pusey, B. J., & Hughes, J. M. (2011). Landscape genetic analysis of the tropical freshwater fish *Mogurnda mogurnda* (Eleotridae) in a monsoonal river basin: Importance of hydrographic factors and population history. *Freshwater Biology*, *56*(5), 812–827. https://doi.org/10.1111/j.1365-2427.2010.02527.x

Cowen, R. K., & Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, *1*, 443–466. https://doi.org/10.1146/annurev.marine.010908.163757

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dawson, M. (2012). Species richness, habitable volume, and species densities in freshwater, the sea, and on land. *Frontiers of Biogeography*, *4*, 105–116. https://doi.org/10.21425/F54312675.

Dias, M. S., Cornu, J.-F., Oberdorff, T., Lasso, C. A., & Tedesco, P. A. (2013). Natural fragmentation in river networks as a driver of speciation for freshwater fishes. *Ecography*, *36*(6), 683–689. https://doi.org/10.1111/j.1600-0587.2012.07724.x

Dionne, M., Caron, F., Dodson, J. J., & Bernatchez, L. (2008). Landscape genetics and hierarchical genetic structure in Atlantic salmon: The interaction of gene flow and local adaptation. *Molecular Ecology*, *17*(10), 2382–2396. https://doi.org/10.1111/j.1365-294X.2008.03771.x

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Elmer, K. R., Fan, S., Kusche, H., Luise Spreitzer, M., Kautt, A. F., Franchini, P., & Meyer, A. (2014). Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature Communications*, *5*, 5168. https://doi.org/10.1038/ncomms6168

Ford, A. G. P., Bullen, T. R., Pang, L., Genner, M. J., Bills, R., Flouri, T., Ngatunga, B. P., Rüber, L., Schliewen, U. K., Seehausen, O., Shechonge, A., Stiassny, M. L. J., Turner, G. F., & Day, J. J. (2019). Molecular phylogeny of Oreochromis (Cichlidae: Oreochromini) reveals mito-nuclear discordance and multiple colonisation of adverse aquatic environments. *Molecular Phylogenetics and Evolution*, *136*, 215–226. https://doi.org/10.1016/j.ympev.2019.04.008

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973–983. https://doi.org/10.1534/genetics.113.160572

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, *43*(10), 1031–1034. https://doi.org/10.1038/ng.937

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*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Hansen, M. M., Limborg, M. T., Ferchaud, A. L., & Pujolar, J. M. (2014). The effects of Medieval dams on genetic divergence and demographic history in brown trout populations. *BMC Evolutionary Biology*, *14*, 122. https://doi.org/10.1186/1471-2148-14-122

Harrison, I. J., Brummett, R., & Stiassny, M. L. J. (2016). The Congo River Basin. In C. M. Finlayson, G. R. Milton, R. C. Prentice, & N. C. Davidson (Eds.), *The Wetland book: II: Distribution, description and conservation* (pp. 1–18). Springer Netherlands. https://doi.org/10.1007/978-94-007-6173-5_92-1

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016). RADcap: Sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, *16*(5), 1264–1278. https://doi.org/10.1111/1755-0998.12566

Horn, M. H. (1972). The amount of space available for marine and freshwater fishes. *Fishery Bulletin*, *70*, 1295–1297.

Jackson, P. R., Oberg, K. A., Gardiner, N., & Shelton, J. (2009). *Velocity mapping in the lower Congo River: A first look at the unique bathymetry and hydrodynamics of Bulu Reach, West Central Africa*. In: Proceedings of the International Association for Hydraulic Research Congress.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kanno, Y., Vokoun, J. C., & Letcher, B. H. (2011). Fine-scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks. *Molecular Ecology*, *20*(18), 3711–3729. https://doi.org/10.1111/j.1365-294X.2011.05210.x

Kautt, A. F., Machado-Schiaffino, G., & Meyer, A. (2016). Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of Nicaraguan crater lake cichlids. *PLOS Genetics*, *12*(6), e1006157. https://doi.org/10.1371/journal.pgen.1006157

Kling, M. M., & Ackerly, D. D. (2021). Global wind patterns shape genetic differentiation, asymmetric gene flow, and genetic diversity in trees. *Proceedings of the National Academy of Sciences*, *118*(17), e2017317118. https://doi.org/10.1073/pnas.2017317118

Knowlton, N., & Weigt, L. A. (1998). New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1412), 2257–2263. https://doi.org/10.1098/rspb.1998.0568

Kocher, T. D. (2004). Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Reviews Genetics*, *5*, 288–298. https://doi.org/10.1038/nrg1316

Lepais, O., & Weir, J. T. (2014). SimRAD: An R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Molecular Ecology Resources*, *14*(6), 1314–1321. https://doi.org/10.1111/1755-0998.12273

Lévêque, C., Oberdorff, T., Paugy, D., Stiassny, M. L. J., & Tedesco, P. A. (2008). Global diversity of fish (Pisces) in freshwater. *Hydrobiologia*, *595*(1), 545–567. https://doi.org/10.1007/s10750-007-9034-0

Levin, B. A., Simonov, E., Dgebuadze, Y. Y., Levina, M., & Golubtsov, A. S. (2020). In the rivers: Multiple adaptive radiations of cyprinid fishes (*Labeobarbus*) in Ethiopian Highlands. *Scientific Reports*, *10*, 7192. https://doi.org/10.1038/s41598-020-64350-4

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*(S1), 52–70. https://doi.org/10.3109/19401736.2010.537748

Lujan, N. K., Weir, J. T., Noonan, B. P., Lovejoy, N. R., & Mandrak, N. E. (2020). Is Niagara Falls a barrier to gene flow in riverine fishes? A test using genome-wide SNP data from seven native species. *Molecular Ecology*, *29*(7), 1235–1249. https://doi.org/10.1111/mec.15406

Lundberg, J. G., Kottelat, M., Smith, G. R., Stiassny, M. L. J., & Gill, A. C. (2000). So many fishes, so little time: An overview of recent ichthyological discovery in continental waters. *Annals of the Missouri Botanical Garden*, *87*(1), 26–62. https://doi.org/10.2307/2666207

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77. https://doi.org/10.1111/1755-0998.12592

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., Miska, E. A., Durbin, R., Genner, M. J., & Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science (New York, N.Y.)*, *350*(6267), 1493–1498. https://doi.org/10.1126/science.aac9927

Markert, J. A., Schelly, R. C., & Stiassny, M. L. (2010). Genetic isolation and morphological divergence mediated by high-energy rapids in two cichlid genera from the lower Congo rapids. *BMC Evolutionary Biology*, *10*, 149. https://doi.org/10.1186/1471-2148-10-149

Maruyama, T. (1969). Genetic correlation in the stepping stone model with non-symmetrical migration rates. *Journal of Applied Probability*, *6*(3), 463–477. https://doi.org/10.2307/3212095

Miller, E. C. (2021). Comparing diversification rates in lakes, rivers, and the sea. *Evolution*, *75*(8), 2055–2073. https://doi.org/10.1111/evo.14295

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Morrissey, M. B., & de Kerckhove, D. T. (2009). The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. *The American Naturalist*, *174*(6), 875–889. https://doi.org/10.1086/648311

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Oberg, K., Shelton, J. M., Gardiner, N., & Jackson, P. R. (2008). *Discharge and other hydraulic measurements for characterizing the hydraulics of lower Congo River*. In: Proceedings of the International Association for Hydraulic Research Congress.

Ochoa, L. E., Pereira, L. H. G., Costa-Silva, G. J., Roxo, F. F., Batista, J. S., Formiga, K., Foresti, F., & Oliveira, C. (2015). Genetic structure and historical diversification of catfish *Brachyplatystoma platynemum* (Siluriformes: Pimelodidae) in the Amazon basin with implications for its conservation. *Ecology and Evolution*, *5*(10), 2005–2020. https://doi.org/10.1002/ece3.1486

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*, 94–100. https://doi.org/10.1038/ng.3464

Poelstra, J. W., Richards, E. J., & Martin, C. H. (2018). Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. *Molecular Ecology*, *27*(21), 4270–4288. https://doi.org/10.1111/mec.14784

Pringle, J. M., Blakeslee, A. M. H., Byers, J. E., & Roman, J. (2011). Asymmetric dispersal allows an upstream region to control population structure throughout a species' range. *Proceedings of the National Academy of Sciences*, *108*(37), 15288–15293. https://doi.org/10.1073/pnas.1100473108

'Puckett, E. (2018). *Making conStruct input files*. Retrieved from: https://puckettresearch.org/2018/05/02/making-construct-input-files/

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). *Tracer 1.6*. Retrieved from: http://beast.bio.ed.ac.uk/Tracer.

Ridgeway, J. (2016). *Object-oriented approaches to mapping channel habitat on the lower Congo River*. Doctoral dissertation, University of Georgia.

Robert, M. (1946). *Le Congo physique* (3rd ed.). Presse Universitaires de France, Liège.

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.

Rocha, L. A., & Bowen, B. W. (2008). Speciation in coral-reef fishes. *Journal of Fish Biology*, *72*(5), 1101–1121. https://doi.org/10.1111/j.1095-8649.2007.01770.x

Runge, J. (2007). The Congo River, Central Africa. In A. Gupta (Ed.), *Large rivers: Geomorphology and management* (pp. 293–309). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470723722.ch14

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

Samarasin, P., Shuter, B. J., & Rodd, F. H. (2017). After 100 years: Hydroelectric dam-induced life-history divergence and population genetic changes in sockeye salmon (*Oncorhynchus nerka*). *Conservation Genetics*, *18*, 1449–1462. https://doi.org/10.1007/s10592-017-0992-0

Santos, M. E., & Salzburger, W. (2012). How cichlids diversify. *Science*, *338*(6107), 619–621. https://doi.org/10.1126/science.1224818

Schelly, R. C., & Stiassny, M. L. J. (2004). Revision of the Congo River *Lamprologus* Schilthuis, 1891 (Teleostei: Cichlidae), with descriptions of two new species. *American Museum Novitates*, *3451*, 40. https://doi.org/10.1206/0003-0082(2004)451<0001:ROTCRL>2.0.CO;2

Schobert, C. S., Stiassny, M. L. J., Schwab, I. R., Zeiss, C., Schelly, R. C., & Dubielzig, R. R. (2013). Comparative ocular anatomy in a blind African cichlid fish, *Lamprologus lethops*. *Veterinary Ophthalmology*, *16*(5), 359–364. https://doi.org/10.1111/vop.12006

Schwarzer, J., Misof, B., Ifuta, S. N., & Schliewen, U. K. (2011). Time and origin of cichlid colonization of the lower Congo rapids. *PLoS One*, *6*(7), e22380. https://doi.org/10.1371/journal.pone.0022380

Stiassny, M. (1997). A phylogenetic overview of the lamprologine cichlids of Africa (Teleostei, Cichlidae): A morphological perspective. *South African Journal of Science*, *93*, 513–523.

Stiassny, M. L. J., & Alter, S. E. (2021). Evolution in the fast lane: Diversity, ecology, and speciation of cichlids in the lower Congo River. In M. E. Abate, & D. L. G. Noakes (Eds.), *The behavior, ecology and evolution of cichlid fishes* (pp. 107–133). https://doi.org/10.1007/978-94-024-2080-7_4

Sullivan, J. P., Lavoué, S., & Hopkins, C. D. (2002). Discovery and phylogenetic analysis of a riverine species flock of African electric fishes (Mormyridae: Teleostei). *Evolution*, *56*(3), 597–616. https://doi.org/10.1111/j.0014-3820.2002.tb01370.x

Svardal, H., Salzburger, W., & Malinsky, M. (2021). Genetic variation and hybridization in evolutionary radiations of cichlid fishes. *Annual Review of Animal Biosciences*, *9*, 55–79. https://doi.org/10.1146/annurev-animal-061220-023129

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*(2), 437–460. https://doi.org/10.1093/genetics/105.2.437

Tedesco, P. A., Paradis, E., Lévêque, C., & Hugueny, B. (2017). Explaining global-scale diversification patterns in actinopterygian fishes. *Journal of Biogeography*, *44*(4), 773–783. https://doi.org/10.1111/jbi.12905

Thomaz, A. T., Christie, M. R., & Knowles, L. L. (2016). The architecture of river networks can drive the evolutionary dynamics of aquatic populations. *Evolution*, *70*(3), 731–739. https://doi.org/10.1111/evo.12883

Tin, M. M. Y., Rheindt, F. E., Cros, E., & Mikheyev, A. S. (2015). Degenerate adaptor sequences for detecting PCR duplicates in reduced representation sequencing data improve genotype calling accuracy. *Molecular Ecology Resources*, *15*(2), 329–336. https://doi.org/10.1111/1755-0998.12314

Tishechkin, D. Y. (2020). The role of sympatric and allopatric speciation in the origin of biodiversity of herbivorous insects, with palaearctic Species of the Genus Macropsis Lewis, 1836 taken as an example (Homoptera, Auchenorrhyncha, Cicadellidae, Eurymelinae, Macropsini). *Entomological Review*, *100*(8), 1039–1064. https://doi.org/10.1134/S0013873820080011

Tobler, M., DeWitt, T. J., Schlupp, I., García de León, F. J., Herrmann, R., Feulner, P. G. D., Tiedemann, R., & Plath, M. (2008). Toxic hydrogen sulfide and dark caves: Phenotypic and genetic divergence across two abiotic environmental gradients in *Poecilia mexicana*. *Evolution*, *62*(10), 2643–2659. https://doi.org/10.1111/j.1558-5646.2008.00466.x

Tornabene, L., Valdez, S., Erdmann, M., & Pezold, F. (2015). Support for a 'Center of Origin' in the Coral Triangle: Cryptic diversity, recent speciation, and local endemism in a diverse lineage of reef fishes (Gobiidae: Eviota). *Molecular Phylogenetics and Evolution*, *82*, 200–210. https://doi.org/10.1016/j.ympev.2014.09.012

Torterotot, J. B., Perrier, C., Bergeron, N. E., & Bernatchez, L. (2014). Influence of forest road culverts and waterfalls on the fine-scale distribution of brook trout genetic diversity in a boreal watershed. *Transactions of the American Fisheries Society*, *143*(6), 1577–1591. https://doi.org/10.1080/00028487.2014.952449

Tougas, S., & Stiassny, M. L. J. (2014). *Lamprologus markerti*, a new lamprologine cichlid (Teleostei: Cichlidae) endemic to the lower Congo River in the Democratic Republic of Congo, west-central Africa. *Zootaxa*, *3852*(3), 391–400. https://doi.org/10.11646/zootaxa.3852.3.8.

Tripp, E. A., Tsai, Y. E., Zhuang, Y., & Dexter, K. G. (2017). RADseq dataset with 90% missing data fully resolves recent radiation of Petalidium (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution*, *7*(19), 7920–7936. https://doi.org/10.1002/ece3.3274

Vavrek, M. J. (2011). Fossil: Palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica*, *14*(1), 16.

Vega-Retter, C., Muñoz-Rojas, P., Rojas-Hernández, N., Copaja, S., Flores-Prado, L., & Véliz, D. (2020). Dammed river: Short- and long-term consequences for fish species inhabiting a river in a Mediterranean climate in central Chile. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *30*(12), 2254–2268. https://doi.org/10.1002/aqc.3425

Washburn, B. A., Cashner, M. F., & Blanton, R. E. (2020). Small fish, large river: Surprisingly minimal genetic structure in a dispersal-limited, habitat specialist fish. *Ecology and Evolution*, *10*(4), 2253–2268. https://doi.org/10.1002/ece3.6064

Winker, K. (2021). An overview of speciation and species limits in birds. *Ornithology*, *138*(2), 1–27. https://doi.org/10.1093/ornithology/ukab006

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, *16*(2), 97–159. https://doi.org/10.1093/genetics/16.2.97

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138. https://doi.org/10.1093/genetics/28.2.114

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Kurata, N. P., Hickerson, M. J., Hoffberg, S. L., Gardiner, N., Stiassny, M. L. J., & Alter, S. E. (2022). Riverscape genomics of cichlid fishes in the lower Congo: Uncovering mechanisms of diversification in an extreme hydrological regime. *Molecular Ecology*, *00*, 1–17. https://doi.org/10.1111/mec.16495