Received: 24 January 2020 | Revised: 19 November 2020 | Accepted: 28 November 2020

DOI: 10.1111/jai.14167

ORIGINAL ARTICLE

Journal of Applied Ichthyology | WILEY

Genetic diversity and population structure of selected lacustrine and riverine populations of African catfish, *Clarias gariepinus* (Burchell, 1822), in Kenya

George W. Alal[1] | James E. Barasa[1] | Emily J. Chemoiwa[2] | Boaz Kaunda-Arara[1] | Peter Akoll[3] | Charles Masembe[3]

[1]Department of Fisheries and Aquatic Sciences, University of Eldoret, Eldoret, Kenya

[2]Department of Biological Sciences, University of Eldoret, Eldoret, Kenya

[3]Department of Zoology, Entomology and Fisheries Sciences, College of Natural Sciences, Makerere University, Kampala, Uganda

Correspondence
James E. Barasa, Department of Fisheries and Aquatic Sciences, University of Eldoret, P.O Box 1125-30100, Eldoret, Kenya.
Email: barkanoti@gmail.com

Funding information
National Commission for Science, Technology and Innovation; National Research Fund, Kenya; Austrian Partnership Programme in Higher Education and Research for Development (APPEAR) STRECAFISH Project; National Research Fund

Abstract

Determining the genetic characteristics of natural fish stocks is useful for conservation and aquaculture programs. For African catfish, *Clarias gariepinus*, genetic characterization could help identify populations suitable as brood stock for culture, and those in need of conservation. This study determined the genetic diversity, population structure, and demographic history of *C. gariepinus* from Lakes Victoria (LV), Kenyatta (LKE), Kamnarok (LKA), and Rivers Nyando (NR), Tana (TR) and Sosiani (SR) in Kenya. Using 128 DNA sequences of D-loop control region, 34 haplotypes were recovered, of which 79.4% were singletons. Only 7 haplotypes were shared between sites, implying little gene flow between sites. Number of haplotypes was highest in LKE and NR populations and lowest in SR. Haplotype diversity was highest in LV, and lowest in SR, while, nucleotide diversity was highest in LKA and lowest in LV. Phylogenetic analyses revealed five clusters: Lakes Victoria, Kamnarok and Kenyatta, and Rivers Tana and Nyando, from both maximum likelihood tree and minimum spanning network. This, together with significant F_{ST} values among the sites imply population differentiation. Mismatch distributions were multi-modal in LKA, LKE, NR and TR, signifying demographic equilibria. Neutrality tests Tajima`s D values for the sampled populations were negative and significantly different, suggesting stable populations. These results show the existence of genetically distinct populations of *C. gariepinus* that require spatially explicit management actions such as reducing fishing pressure, pollution, minimizing habitat destruction and fragmentation for sustainable utilisation of stocks.

KEYWORDS

Clarias gariepinus, genetic differentiation, haplotypes, mitochondrial DNA control region, multi-modal distributions

1 | INTRODUCTION

The African sharp-tooth catfish, *Clarias gariepinus* (Burchell 1822), has a Pan-African distribution, ranging from the Nile River Basin to West Africa and from Algeria to Southern Africa (Cambray, 2003). *Clarias gariepinus* inhabits both riverine and lacustrine environments, apart from streams and swamps, where it plays an important ecological role as a predator (Corbet, 1961). The species grows to total lengths of up to 130 cm and weighs over 30 kg (Bruton, 1979; Cambray, 2003). Its wide distribution is partly due to an omnivorous

feeding habit and ability to tolerate a wide range of environmental variables (Bruton, 1979). In this regard, the species is especially successful in large rivers, where it swims in inundated areas for breeding, but also inhabits different habitats of varying sizes and conditions. The species has high fecundity (Owiti & Dadzie, 1989) and fast growth rates (Bruton, 1979), which make it an excellent aquaculture species, cultured for food and nutrition security and livelihoods.

In East Africa, juveniles of *Clarias* are also used as live bait for Nile perch, *Lates niloticus* using long line and hooks in Lake Victoria (Chitamwebwa et al., 2009; Ngugi et al., 2005). In this regard, there is indiscriminate collection of *Clarias* juveniles from natural habitats for use as live bait by fishermen, a practice which increases fishing pressure on natural populations, endangers the fishermen's health and often fails to meet the numbers of live bait required. In order to reduce this overexploitation, artificial propagation of *C. gariepinus* at hatcheries has been recommended (Kaufman & Ochumba, 1993), as a source of *Clarias* juveniles, which also generates income and livelihood for farmers (Barasa et al., 2017). However, the efficiency of artificial propagation of the species at hatcheries is reduced by poor survival of *C. gariepinus* larvae (Sulem et al., 2006). Recent studies to address this challenge have focused on genetic characterization of *C. gariepinus* populations in order to identify suitable populations for use as brood stock (Barasa et al., 2014, 2016; 2017; Ojiambo, 2015) and those populations in need of conservation (Barasa et al., 2017).

Preservation of genetic diversity is an important goal of conservation programmes, since genetic diversity influences adaptation and persistence of a species in the environment (Lande, 1988). Furthermore, for *C. gariepinus*, genetic diversity correlates with some fitness traits (Barasa, 2018), and therefore identifying populations of high genetic diversity is crucial in the sustainable utilization of the species. As a follow up to previous studies on the genetic characterization of *C. gariepinus* in Kenya, this study increased the geographical coverage of *C. gariepinus* populations, for improved culture and conservation of the species in East Africa. The study determined genetic diversity and population structure of *C. gariepinus* from selected lacustrine and riverine habitats in Kenya, using mitochondrial DNA of the D-loop control region.

2 | MATERIALS AND METHODS

2.1 | Fish samples

Samples of *C. gariepinus* were collected from six different sites in Kenya from August, 2016 to May, 2017: three lacustrine sites including Lakes Victoria (LV), Kamnarok (LKA) and Kenyatta (LKE), and three riverine sites including Rivers Nyando (NR), Sosiani (SR), and Tana (TR)) (Table 1 and Figure 1). The six sites were selected based on considerations such as habitat size, climatic and hydrographic conditions, and the fact that all the sites constituted important *C. gariepinus* fishery, exploited for food and as source of brood stock for artificial propagation at hatcheries. Sampling of *C. gariepinus* from the sites therefore provided an extensive coverage of the geographical spread of the species in Kenya, in addition to the fact that the sites have varying types and levels of anthropogenic influences on the fishery. It was therefore possible to decipher the effect of these anthropogenic factors on genetic diversity and population structure of *C. gariepinus* in different habitats in Kenya. The sampling station(s) within sites were selected randomly and fish collected using a combination of gill netting, hook and line, and traps and seine nets. Fish samples were identified using a field identification guide (Witte & van Densen, 1995). Pectoral fins were clipped from 128 individuals of *C. gariepinus* (Table 1) as source of DNA using sterilized surgical blade and immediately preserved in 95% ethanol in clean labelled cryovial tubes and kept in the freezer at −20°C before DNA extraction.

2.2 | DNA extraction

Fin clips were thawed, macerated, tissue lysed and incubated at 40°C overnight for digestion in a shaking water bath in preparation for DNA extraction. Total genomic DNA was extracted from approximately 25 mg of fin clip tissue, using the Qiagen DNeasy Tissue Kit (Qiagen GmbH, Germany); following the manufacturer`s protocol with minor modifications. Centrifugation for spinning digested content was set at 10,000 revolutions per minute (rpm) except for the final, which was done at 14,000 rpm. Elution was done with 150 μl of AE elution buffer repeated twice for maximum yield. Presence and quality of the extracted genomic DNA were visualized using 2% agarose gel electrophoresis. The DNA was stored at −20°C prior to further analysis.

2.3 | PCR amplification

The hypervariable mitochondrial DNA control region (D-loop) was Polymerase Chain Reaction (PCR) amplified in a thermal cycler (ABI 9700) using the following primers: forward primer LN20 (5′-ACCACTAGCACCCAAAGCTA-3′) and reverse primer HN20 (5′-GTGTTATGCTTTAGTTAAGC-3′) (Benatchez & Danzmann, 1993). PCR reactions were carried out in a 25 μl volume microcentrifuge vials. The mixtures contained 0.5 μl of 10 pmol of each primer (HN20 and LN20), 12.5 μl of AmpliTaq® Gold 360 Master Mix (Applied Biosystems, USA), 9.5 μl of double distilled water and 2 μl of genomic DNA template.

The PCR thermal cycling conditions were as follows: initial denaturation at 95°C for 1 min, 35 cycles of denaturation at 94°C for 30 s, primer annealing at 49.3°C for 1 min, initial extension at 72°C for 50 s, followed by a final extension of 72°C for 10 min. All PCR reactions were performed in a Thermal Mastercycler (Eppendorf, Germany) PCR system 9700. In every PCR, a DNA free template was included as a negative control. The PCR products were visualized on gel electrophoresis (100v, 30 min) in 2% agarose gel, and size compared against

lambda DNA ladder. The Amplicons were purified using QIAquick PCR purification kit (Qiagen GmbH, Germany) following manufacturer`s instructions. The products were visualized in 2% Nusieve agarose gel stained. The purified products were cycle-sequenced and analysed on ABI 3,100 automated-sequencer. One hundred and twenty-eight PCR products were selected with correct band, good quality and sent to Macrogen Laboratory (Netherlands) for sequencing.

2.4 | Data analysis

The chromatograms for the forward and reverse DNA strands were refined and consensus sequences generated from contigs in the Sequencher program v5.4.6 (Gene Codes, 2016). The consensus sequences were edited in BioEdit v7.1.11 (Hall, 2005) and aligned using the complete alignment application tool in Clustal X v 2.0 (Larkin et al., 2007). The sequences were compared with nucleotide sequences in the GenBank using the Basic Local Alignment Search Tool (BLAST) to confirm species identity. Thereafter, the aligned sequences were used for phylogenetic and genetic diversity analyses in MEGA X (Kumar et al., 2018).

Genetic diversity within *C. gariepinus* populations was quantified as the number of haplotypes, haplotype diversity (*h*), nucleotide diversity (π), number of singletons (with percentages), number of shared haplotypes (with percentages), polymorphic sites in DnaSP v6.11.01 (Rozas et al., 2017). Consequently, the distribution and identification of shared haplotypes was performed in DnaSP, according to segregating sites. JModelTest v2.1.10 (Darriba et al., 2012) was used to determine the most likely model of evolution for the mtDNA sequences. Hierarchical distribution of genetic structure of *C. gariepinus* from the six sites was analysed by standard analysis of molecular variance (AMOVA) computations (Excoffier, 2004) using haplotype pairwise differences performed to partition genetic variation in Arlequin v3.5 (Excoffier & Lischer, 2010). The AMOVA partitioned total variance into covariance components due to within populations, among populations and among group variances. The covariance computed fixation indices included F_{SC}, F_{ST} and, F_{CT} respectively as defined by Wright (1965).

Phylogenetic relationships among lacustrine and riverine populations of *C. gariepinus* mtDNA control region sequence haplotypes were calculated using Maximum likelihood (ML) methods. The best-fit model for DNA evolution was General Time Reversible (GTR + G) selected by Akaike Information Criterion (AIC) using JModelTest v2.1.10 (Darriba et al., 2012). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the ML approach, and then selecting the topology with superior log-likelihood value (Darriba et al., 2012) performed in PAUP* v4.0a167 (Swofford, 2002). The tree was visualized in Interactive tree of life (iTOL) (Letunic and Bork, 2016).

Minimum spanning network showing the phylogenetic relationships between haplotypes of *C. gariepinus* from lacustrine populations (LKA, LKE and LV) and riverine populations (NR, TR and SR) was drawn using MEGA X (Kumar et al., 2018) borrowing labels from the ML tree, with a median-joining approach. The circle size is proportional to the haplotype frequency. Moreover, haplotypes were identified by the number of each haplotype and their resultant branch length represent the number of mutation steps.

The account of demographic history of *C. gariepinus* population expansion and decline within and between sampling sites was simulated in (coalescent simulations) DnaSP v6.11.01 (Rozas et al., 2017). Additionally, mismatch distribution analysis, neutrality tests; Fu`s Fs (Fu, 1997) and Tajima`s D (Tajima, 1989) were estimated in DnaSP v6.11.01 (Rozas et al., 2017), to discern demographic changes amongst the *C. gariepinus* populations.

3 | RESULTS

3.1 | Genetic diversity of *C. gariepinus* inferred from mtDNA D-loop

Sequences were submitted to the GenBank database and are publicly available under their respective accession numbers (Table 1). Genetic diversity of *C. gariepinus* based on Haplotype diversity (*h*) and Nucleotide diversity (π) was generally higher in lacustrine populations than riverine populations (Table 2). The lakes' and River Nyando fish populations exhibited high *h*, which were not significantly different from each other (LV = 0.911 ± 0.042; LKA = 0.774 ± 0.071 LKE = 0.866 ± 0.059 and NR = 0.830 ± 0.068) compared to SR and TR populations (SR = 0.518 ± 0.122 and TR = 0.620 ± 0.099). Consequently, Nucleotide diversity values were higher in LKA and LKE populations (LKA = 0.118 ± 0.056 and LKE = 0.109 ± 0.059), than NR, SR and TR populations (NR = 0.060 ± 0.041; SR = 0.031 ± 0.027

TABLE 1 Sampling sites, Population codes, coordinates of sampling sites, sample sizes and GenBank accession numbers of sequences for 128 samples of *C. gariepinus* collected from different lacustrine and riverine habitats in Kenya

Sites	Population code	Coordinates	Sample size	Altitude	GenBank Sequence accession numbers
Lake Victoria	LV	1°00′S, 33°00′E	20	1,135m a.s.l	MK014577-MK014596
Lake Kamnarok	LKA	0°37′N, 35°37′E	20	1,058m a.s.l	MK014597-MK014616
Lake Kenyatta	LKE	2°24′S, 40°40′E	23	10m a.s.l	MK014617-MK014639
River Nyando	NR	0°09′S, 34°55′E	23	1,156m a.s.l	MK014640-MK014662
River Sosiani	SR	0°32´N, 35°13´E	23	2,084 a.s.l	MK014663-MK014685
River Tana	TR	1°30′S, 40°0′E	19	1,975m a.s.l	MK014686-MK014704

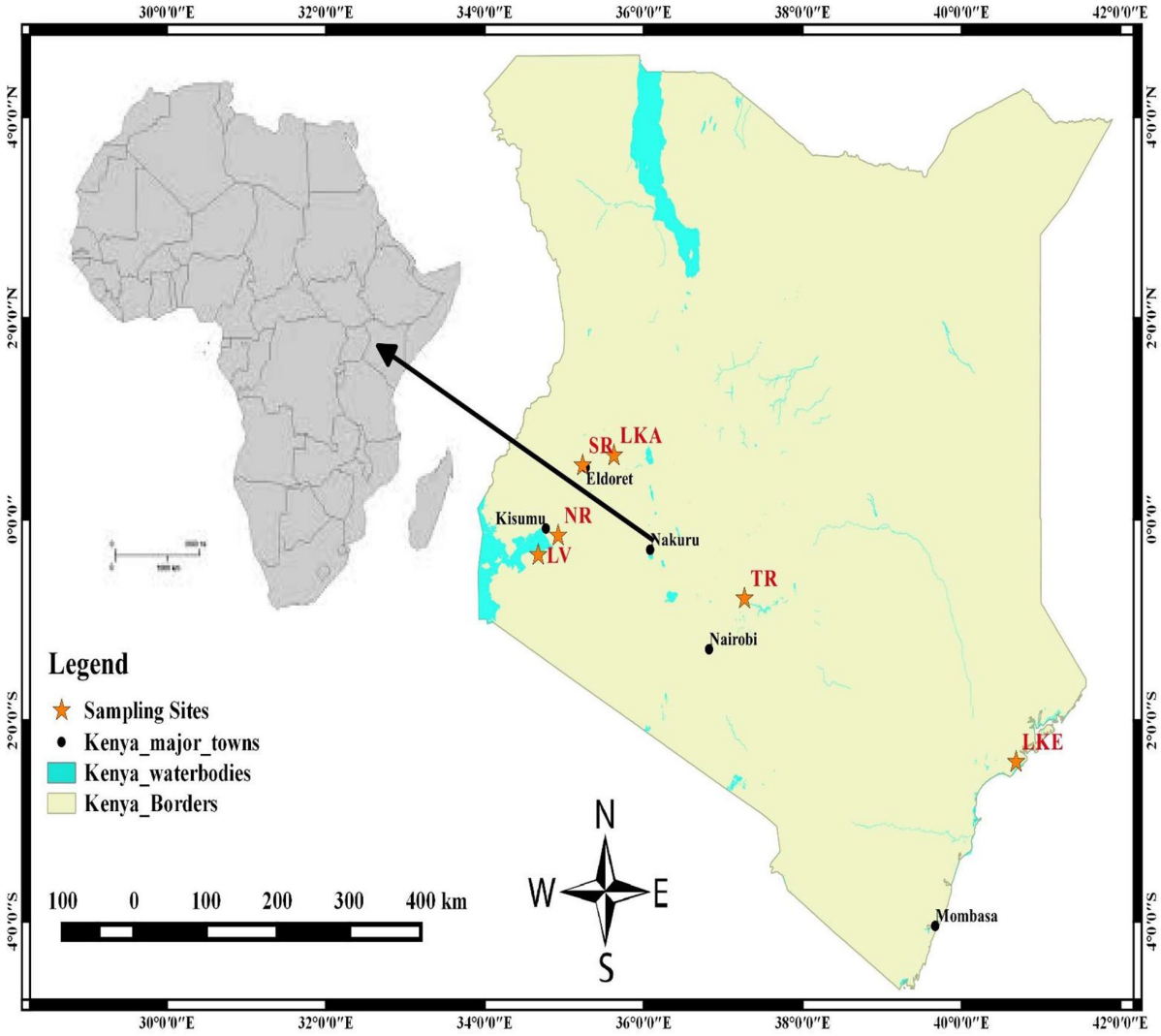

FIGURE 1 Map of field sampling locations for *Clarias gariepinus* from six different sites in Kenya. The three lakes sampled included Lakes Victoria (LV), Kamnarok (LKA), and Kenyatta (LKE). The three rivers included Rivers Nyando (NR), Sosiani (SR) and Tana (TR)

TABLE 2 Genetic diversity estimates of *C. gariepinus* from six aquatic systems in Kenya from 570 bp of mtDNA D-loop control region sequences: LV = Lake Victoria, NR = River Nyando, LKA = Lake Kamnarok, SR = River Sosiani, LKE = Lake Kenyatta and TR = River Tana; $h \pm$ S.D (the haplotype diversity ± Standard Deviation), $\pi \pm$ S.D (the nucleotide diversity ± Standard Deviation), Singletons, shared haplotypes, Theta S (Θ_S), and Theta Pi ($\Theta\pi$). Numbers in parentheses are percentage values of the respective variables

Sites						
Parameters	**LV**	**LKA**	**LKE**	**NR**	**SR**	**TR**
Sample Size	20	20	23	23	23	19
No. of Haplotypes	10(20.41)	8(16.33)	11(22.45)	11(22.45)	4(8.16)	5(10.20)
$h \pm$ S.D	0.911 ± 0.042	0.774 ± 0.071	0.866 ± 0.059	0.830 ± 0.068	0.391 ± 0.125	0.620 ± 0.099
$\pi \pm$ S.D	0.008 ± 0.003	0.118 ± 0.056	0.109 ± 0.059	0.060 ± 0.041	0.031 ± 0.027	0.099 ± 0.056
Singletons	3(11.11)	6(22.22)	9(33.33)	5(18.52)	2(7.41)	2(7.41)
Shared Haplotypes	7(31.82)	2(9.09)	2(9.09)	6(27.27)	2(9.09)	3(13.64)
Polymorphic Sites	24	331	336	277	158	315
Theta S (Θ_S)	8.174	132.761	125.447	87.515	52.292	126.176
Theta Pi ($\Theta\pi$)	0.015	0.261	0.247	0.168	0.097	0.246

and TR = 0.099 ± 0.056). Lake Victoria population had the lowest nucleotide diversity (LV = 0.008 ± 0.003) (Table 2).

In the 570 bp of sequences of mtDNA control region, 423 segregating (polymorphic) sites defined 34 haplotypes. Up to 79.41% (27 out of 34) of the total number of haplotypes occurred as singletons, while 20.59% were shared between populations. Lacustrine and Nyando river populations (LV $n = 10$, LKA $n = 8$, LKE $n = 11$ and NR $n = 11$ respectively) revealed a higher number of haplotypes than SR ($n = 4$) and TR ($n = 5$) (Table 2).

3.2 | Population structure of *C. gariepinus* inferred from mtDNA D-loop

The Maximum Likelihood tree revealed five clades: Lake Victoria, River Nyando, Lake Kamnarok, Lake Kenyatta and River Tana clusters (Figure 2). Most of the samples from SR clustered with samples of LV, with haplotype 1 having the highest number of samples ($n = 45$). Lake Victoria cluster had the highest number of haplotypes and comprised of all the lacustrine and riverine samples. Lake Kamnarok cluster consisted of haplotypes from LKA, LV, LKE, NR and TR, while Lake Kenyatta cluster had samples from LKA, LKE and NR (Figure 2). River Nyando cluster had haplotype 34 as the sole sample delineating this population (Figure 2; Table 3).

Consistent with the maximum likelihood tree (Figure 2), the minimum spanning network for the relationship between haplotypes of *C. gariepinus* had 34 haplotypes and 5 clusters (Lake Victoria, River Nyando, Lake Kamnarok, Lake Kenyatta and River Tana clusters; Figure 3). The most abundant haplotype, from which all other haplotypes radiated was haplotype 1, consisting of 45 samples (Figure 3). A total of 29 mutational steps were observed between the haplotypes. Analysis of molecular variance (AMOVA), revealed significant variation of 17.22% occurring among populations within groups and 86.69% occurred within populations, while non-significant variation of −3.91% was noted among groups (Table 4).

There was a significant genetic differentiation (nine out of fifteen pairwise comparisons) among populations (F_{ST} Pairwise comparisons, Table 4). The LV population was significantly different from LKA, LKE and SR ($p < .05$), while both LKA and LKE were significantly different from riverine populations ($p < .05$) (Table 4).

3.3 | Demographic history of *C. gariepinus* inferred from mtDNA D-loop

Mismatch distribution of observed number of pairwise differences between haplotypes of *C. gariepinus* are presented for individual populations. Lake Victoria (LV) population showed uni-modal shaped distribution (Figure 4). The reconstructed demographic history of Lake Kamnarok (LKA) revealed a multi-modal shape (Figure 5). On the other hand, the reconstructed demographic history of Lake Kenyatta (LKE) *C. gariepinus* population revealed multi-modal

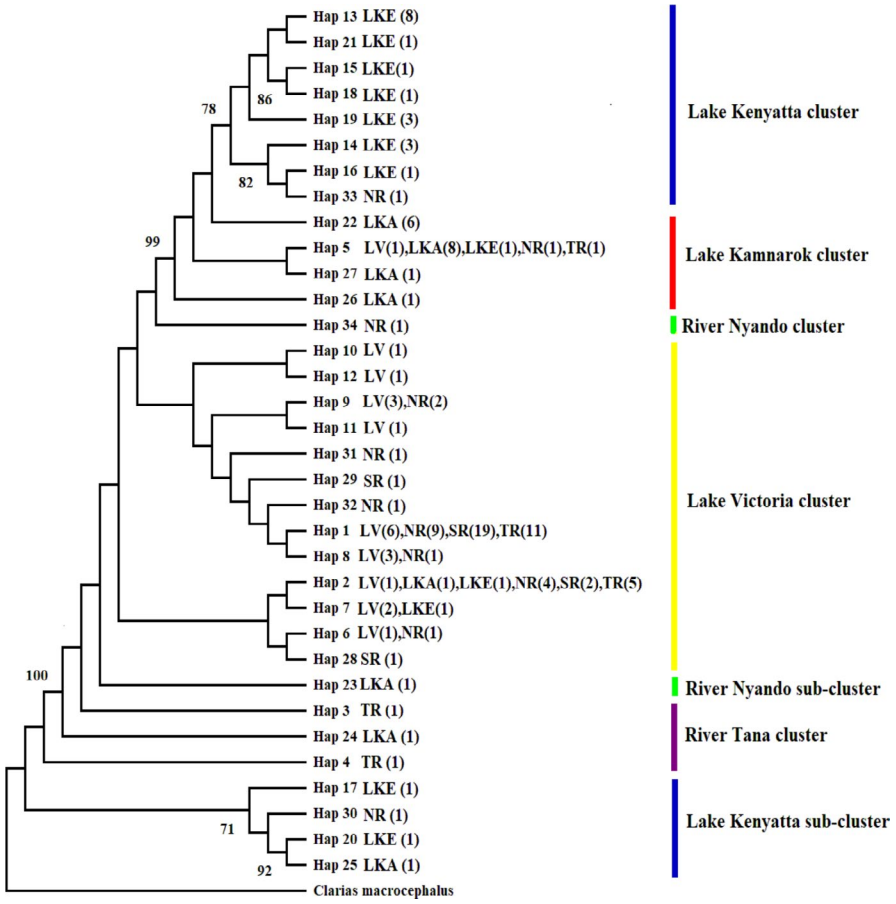

FIGURE 2 Maximum likelihood tree for 34 haplotypes of 128 samples, based on General Time Reversible (GTR + G) model for the Lacustrine and Riverine populations of *C. gariepinus* from six different sampling sites in Kenya. The labelling is same as for the median network (Figure 3). Numbers in parentheses are frequency of individuals in LV, LKA, LKE, NR, SR and TR samples. Numbers on the nodes represent percent bootstrap values, based on 1,000 bootstrap iterations. Bootstrap values below 70% were excluded. *Clarias macrocephalus* was the out-group

TABLE 3 Analysis of Molecular Variance (AMOVA) among Lacustrine and Riverine *C. gariepinus* haplotypes

Source of Variation	*df*	Sum of Squares	Variance Components	% Variation	Fixation Index	*p* - values
Among Groups	2	107.099	−0.723	−3.91	F_{CT} = −0.039	.603^{ns}
Among Populations within groups	3	250.681	3.185	17.22	F_{SC} = 0.166	.001***
Within Populations	122	1956.369	16.036	86.69	F_{ST} = 0.133	.001***
Total	**127**	**2,314.148**	**18.498**			

Note: Not Significant = ns, *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$.

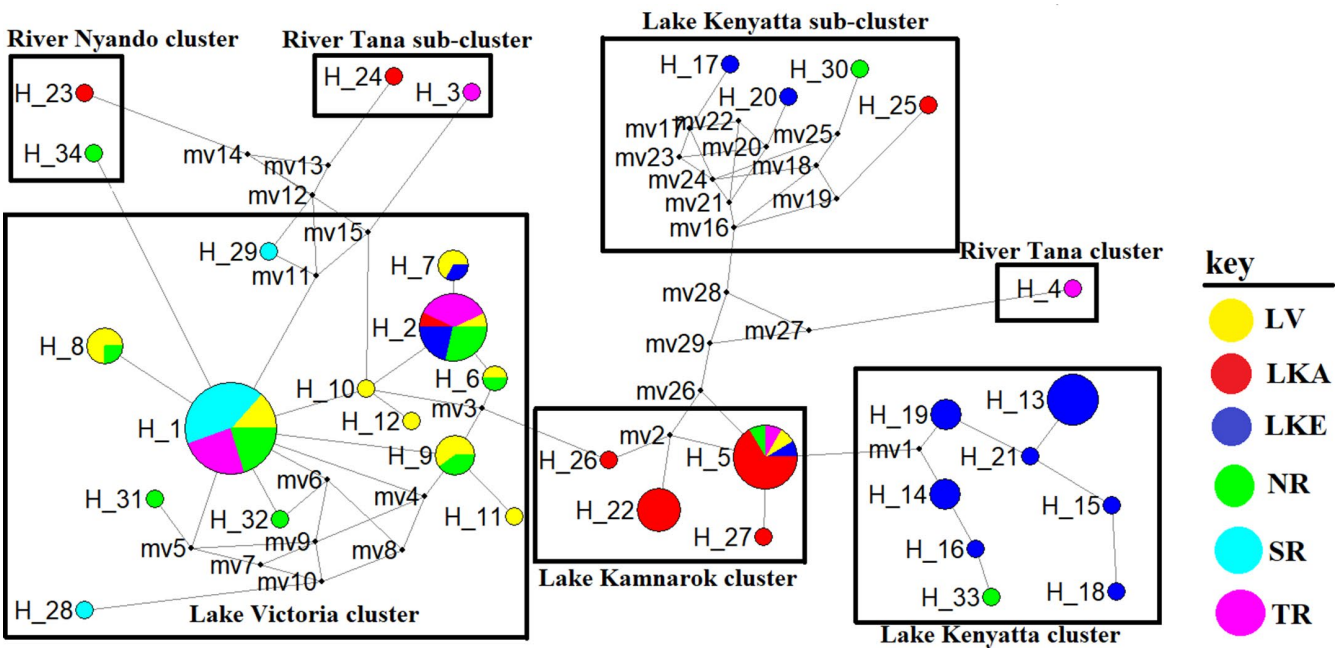

FIGURE 3 Minimum spanning network showing the relationship between haplotypes of *Clarias gariepinus* from lacustrine populations (LKA, LKE and LV) and riverine populations (NR, TR and SR). Size of the circle is proportional to the haplotype frequency. Haplotypes are identified by the number of each haplotype. Branches represent the number of mutational steps

TABLE 4 F_{ST} Pair wise comparisons (below diagonal) and associated P value (above diagonal) of 128 samples of *C. gariepinus* (Burchell, 1822) from six different sampling locations in Kenya, based on 570 bp of mitochondrial D-loop control region sequences

		Associated *p* values					
		LV	LKA	LKE	NR	SR	TR
F_{ST}	LV		.000	.000	.775	.018	.135
	LKA	.251*		.144	.000	.000	.000
	LKE	.285*	.038		.000	.000	.000
	NR	−.008	.152*	.179*		.405	.739
	SR	.011*	.252*	.280*	−.003		.207
	TR	.016	.124*	.152*	−.021	.007	

Note: Values with asterisk are significantly different ($p < .05$).

mismatch distribution pattern (Figure 6). Riverine populations exhibited unimodal shaped mismatch distributions. Nonetheless, a distinct multi-modal shape was revealed in River Nyando (NR) (Figure 7), while River Sosiani (SR) *C. gariepinus* revealed a uni-modal mismatch distribution shape (Figure 8). Finally, River Tana showed a uni-modal mismatch distribution shape (Figure 9).

Among lacustrine populations, LKA and LKE had higher substitutions (471 and 463 respectively) than LV population (29). Riverine populations showed high substitutions with TR and NR carrying higher values (441 and 323 respectively) than SR population (193). In comparison, more substitutions were recorded in fish samples from lacustrine ecosystems (963) than in riverine ecosystems (957) (Table 5).

Raggedness indices in riverine populations were higher in SR and TR, exhibiting values of 0.360 and 0.399, respectively than NR with a value of 0.030. The lacustrine populations had a higher raggedness index in LKA population (0.207) while LV and LKE revealed lower values of 0.042 and 0.038 respectively. Tajima`s D values for all the populations was negative and statistically significant ($p < .05$). Riverine populations had lower values (SR = −2.734, NR = −2.627 and TR = −2.521; $p = .001$) than lacustrine populations (LV = −1.871, LKE = −2.270; $p = .01$ and LKA = −2.315; $p = .05$) (Table 5). Fu`s Fs value for Lake Victoria population was negative and not significant (−2.336, $p > .05$; Table 5), but significant positive for Lakes Kamnarok and Kenyatta (17.413 and 10.566, $p < .001$). Consequently, rivers

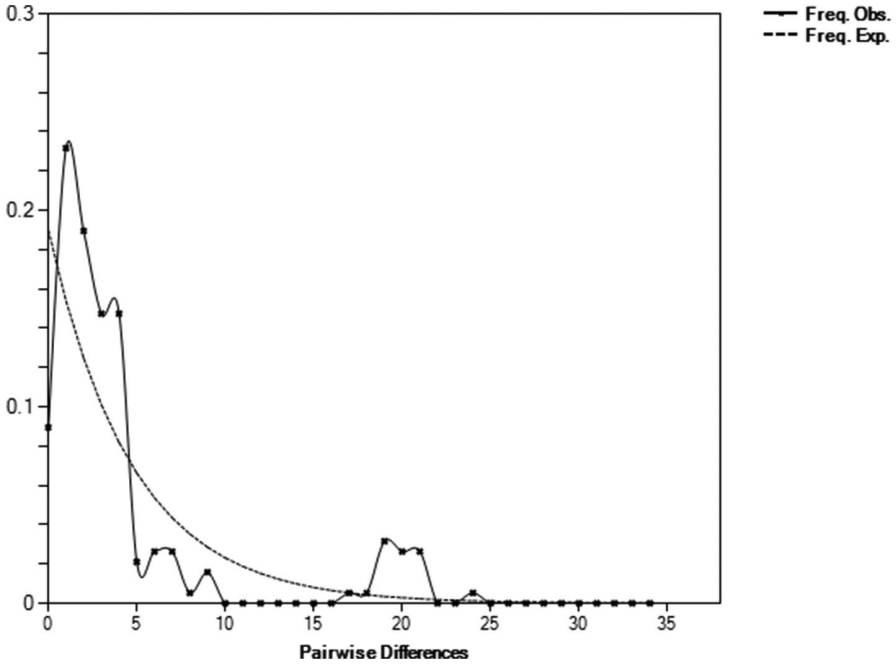

FIGURE 4 The mismatch distribution of individual *Clarias gariepinus* mtDNA control region from the Lake Victoria populations. The X axis shows the number of pairwise differences, and the Y axis shows the frequency of the pairwise comparisons. The plots were based on constant population size change model

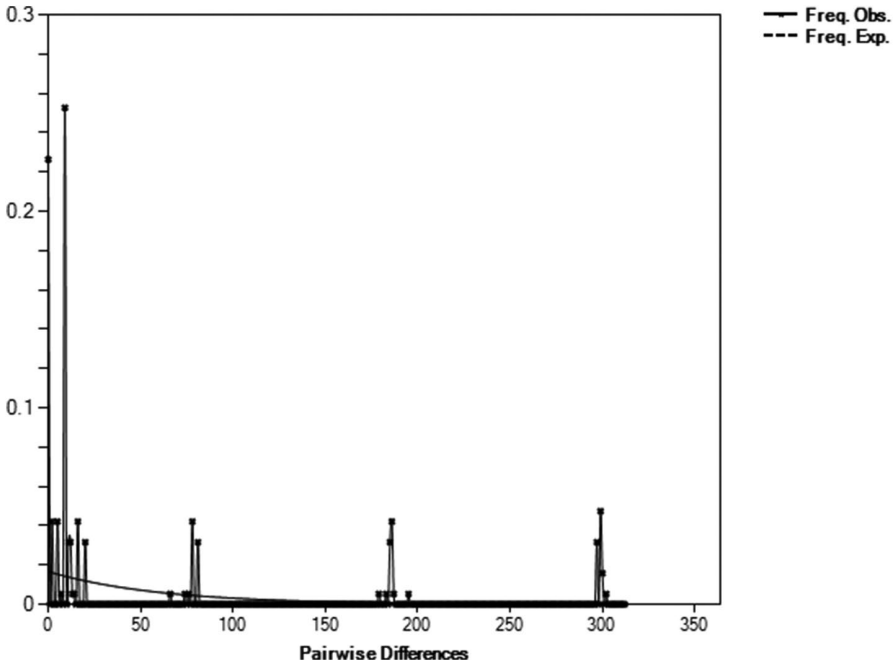

FIGURE 5 The mismatch distribution of individual *Clarias gariepinus* mtDNA control region from the Lake Kamnarok populations. The X axis shows the number of pairwise differences, and the Y axis shows the frequency of the pairwise comparisons. The plots were based on constant population size change model

Nyando, Sosiani and Tana, revealed positive significant values of Fu`s Fs (7.632, $p < .01$; 14.287 and 23.961, $p < .001$) respectively (Table 5).

4 | DISCUSSION

4.1 | Genetic diversity of *C. gariepinus* inferred from the mtDNA D-loop

One way of improving conservation of natural fish populations while simultaneously increasing opportunities for higher aquaculture production is the determination of genetic diversity and population structure of commercially important fish species. Overall, the number of haplotypes and haplotype diversity of *C. gariepinus* reported in this study are comparable to values reported for the species from other lakes in Kenya (Barasa et al., 2017; Nyunja et al., 2017), as well as in the region (Ojiambo, 2015). Although Lake Kenyatta is much smaller than Lake Victoria, the *C. gariepinus* population in Lake Kenyatta had a higher number of haplotypes. This could be attributed to stock augmentation of *C. gariepinus* fishery of Lake Kenyatta by conservation scientists, to boost dwindling fish population due to overfishing and periodic drying of the lake. Indeed, while Lake Kenyatta has been restocked with *C. gariepinus* in the past, these activities were not documented. Previous studies have reported a high number of haplotypes in *C. gariepinus* populations consisting of fish stocks from multiple sources (Barasa et al., 2017; Grobler et al., 1997; Van De

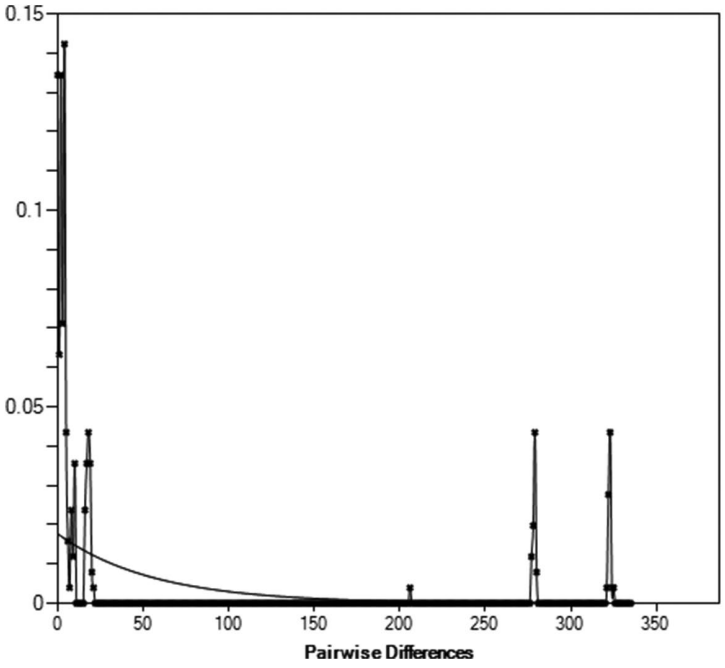

FIGURE 6 The mismatch distribution of individual *Clarias gariepinus* mtDNA control region from the Lake Kenyatta populations. The X axis shows the number of pairwise differences, and the Y axis shows the frequency of the pairwise comparisons. The plots were based on constant population size change model

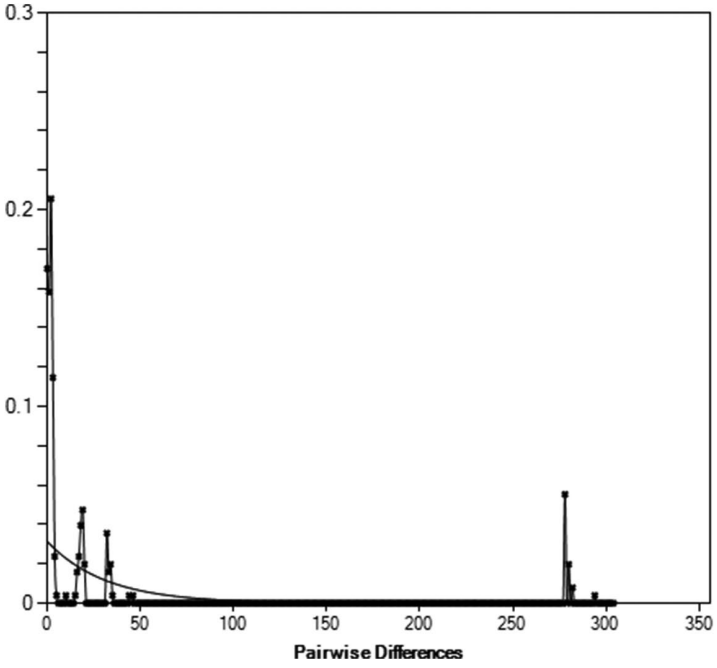

FIGURE 7 The mismatch distribution of individual *Clarias gariepinus* mtDNA control region from the River Nyando populations. The X axis shows the number of pairwise differences, and the Y axis shows the frequency of the pairwise comparisons. The plots were based on constant population size change model

Bank et al., 1992). The higher number of haplotypes in the Nyando River population compared to populations from other rivers may be due to its larger drainage basin of 3,587 km^2 (Gathenya et al., 2011), and the fact that the river flows throughout the year. These characteristics are likely to reduce mortality in the population, and help to maintain higher genetic diversity (Barasa et al., 2016). On the other hand, Tana River which is equally large and perennial had lower number and diversity of haplotypes than Nyando River, and this could be attributed to habitat fragmentation through damming of the river for electricity generation (Vogl et al., 2016). Hydroelectric dams along rivers lead to fish population fragmentation, limiting gene flow (Barocca et al., 2012). In addition, the low genetic diversity indices of the Tana River population could be attributed to overfishing which increases fish mortality, ultimately lowering the number and diversity of haplotypes in *C. gariepinus* (Barasa et al., 2017) and other fishes (Chemoiwa et al., 2013; Muwanika et al., 2012). On the other hand, River Sosiani had the lowest number and diversity of haplotypes among all the sites. This could be attributed to population bottleneck effects (Frankham et al., 2002) associated with small size rivers (Sosiani River is 25 km long), water abstraction and urbanization effects.

The study recorded a high number of singletons for the cumulative sampled *C. gariepinus* populations (27 out of 34) or 79.41%, with only 20.59% carrying shared haplotypes. As the largest water mass

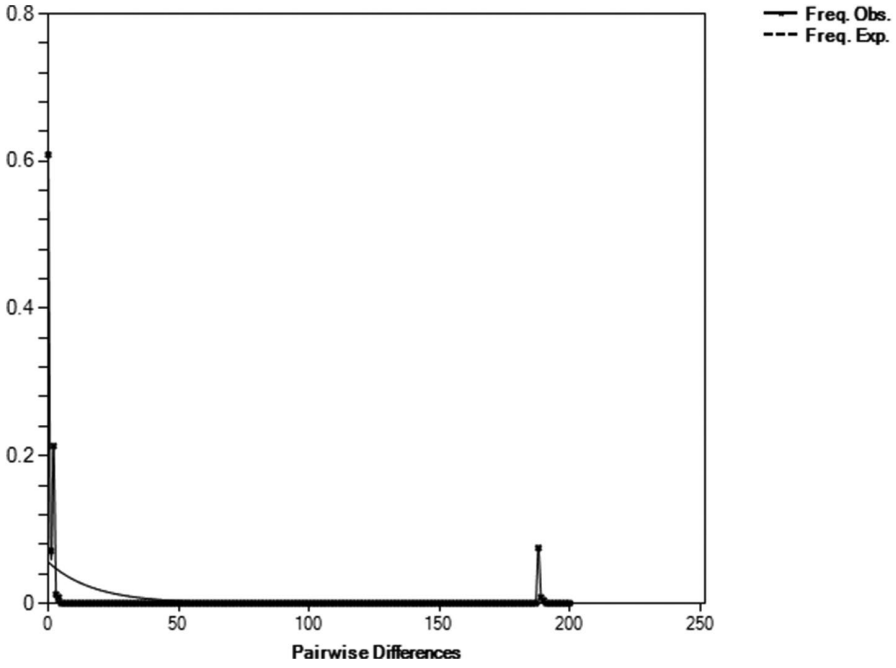

FIGURE 8 The mismatch distribution of individual *Clarias gariepinus* mtDNA control region from the River Sosiani populations. The X axis shows the number of pairwise differences, and the Y axis shows the frequency of the pairwise comparisons. The plots were based on constant population size change model

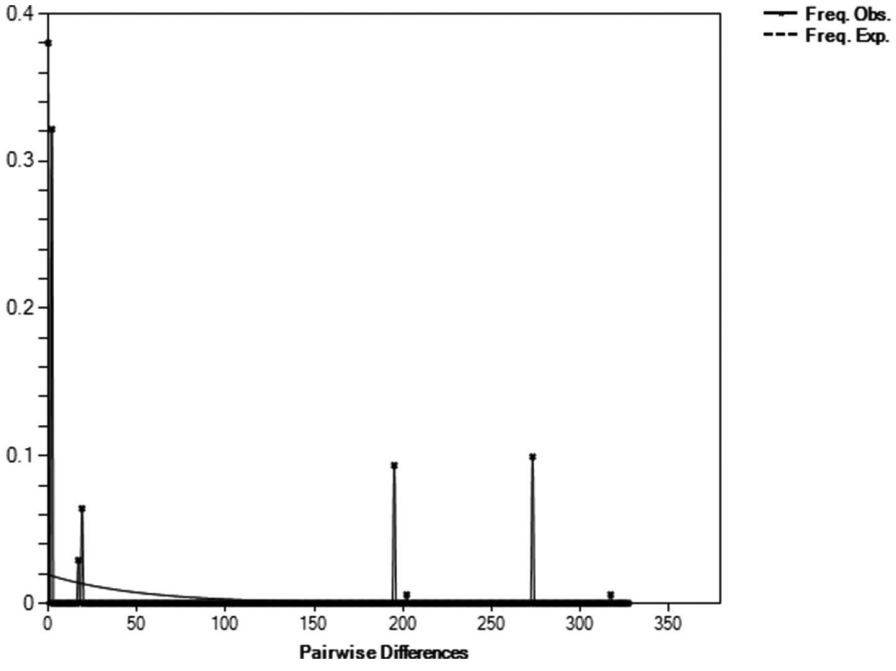

FIGURE 9 The mismatch distribution of individual *Clarias gariepinus* mtDNA control region from the River Tana populations. The X axis shows the number of pairwise differences, and the Y axis shows the frequency of the pairwise comparisons. The plots were based on constant population size change model

of the sampled sites, Lake Victoria possibly harboured a higher number of singletons than any of the populations due to a larger population of *C. gariepinus*. However, the lower number of haplotypes ($n = 3$) compared to LKE ($n = 6$) and LKA ($n = 9$) could be attributed to historical predation by Nile perch, *Lates niloticus*, on *C. gariepinus* in the lake (Njiru et al., 2004; Ogutu-Ohwayo, 1990). Predation by *L. niloticus* is reported to have caused a loss of genetic diversity in *C. gariepinus* of Lake Victoria (Barasa et al., 2014). Meanwhile, the absence of *L. niloticus* in the Nyando River, coupled with the large water mass and perennial nature could have facilitated the emergence and conservation of singletons in this population of *C. gariepinus*. In contrast, small isolated habitats such as LKE and LKA often suffer genetic drift (Frankham et al., 2002), leading to a random emergence of singletons. This, coupled with restocking of Lakes Kenyatta and Kamnarok may account for the higher number of singletons in these populations.

4.2 | Population structure of *C. gariepinus* inferred from the mtDNA D-loop

The Maximum Likelihood tree, Minimum Spanning Network and F_{ST} values showed that *C. gariepinus* populations were differentiated and structured. The differentiation of *C. gariepinus* could be

TABLE 5 Demographic statistics for the combined 570 bp dataset of Lacustrine and Riverine *C. gariepinus* populations. Statistics include mutations, Raggedness, Tajima`s D and Fu`s Fs neutrality tests

Populations	Mutations	SSD	Raggedness	Tajima`s D	Fu`s Fs
Overall Populations	**787**	**0.042^{ns}**	**0.019^{ns}**	**−2.511*****	**12.553*****
Lacustrine	619	0.018^{ns}	0.013^{ns}	−2.346**	10.972***
Riverine	617	0.000^{ns}	0.063^{ns}	−2.780***	17.329***
LV	25	0.010^{ns}	0.060^{ns}	−1.957*	−2.155^{ns}
LKA	465	0.200***	0.205***	−2.340**	17.138***
LKE	481	0.042^{ns}	0.061^{ns}	−2.242**	11.027***
NR	306	0.030^{ns}	0.066^{ns}	−2.646***	7.123**
SR	179	0.210***	0.193***	−2.737***	13.435***
TR	417	0.122^{ns}	0.137***	−2.534***	23.177***

Note: Asterisk indicates significant test (*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$), while ns denotes not significant.

attributed to limited gene flow caused by geographical separation of the sites. Consistently, previous studies recorded significant differentiation among *C. gariepinus* populations from geographically separated lakes in Kenya (Barasa et al., 2017; Nyunja et al., 2017). Indeed, geographical separation and isolation is one of the factors governing genetic divergence among populations (Lande, 1988; Lehmann et al., 1998). These distinct populations therefore constitute important genetic resources that should be conserved by avoiding inter-basin translocation that would result into genetic homogenization (Barasa et al., 2017). The existence of distinct population differentiation in the samples was further supported by results of the analysis of molecular variance (AMOVA), which showed that genetic variation of the *C. gariepinus* samples was mainly accounted for by variation within populations and also because of significant differences among populations.

4.3 | Demographic history of *C. gariepinus* inferred from the mtDNA D-loop

The *C. gariepinus* populations from lakes had multi-modal shaped distributions, suggesting stable populations (Rogers & Harpending, 1992). This distribution conformed to significant value for Tajima's D neutrality test. However, the negative neutrality test suggests that the population expansion of lacustrine *C. gariepinus* is recent implying that the equilibrium exhibited by these populations was achieved after a population expansion. The stability of Lake Victoria *C. gariepinus* population after an expansion could be attributed to the decline in the stock biomass of the predatory *L. niloticus* (Kayanda et al., 2009). Additionally, the exotic water hyacinth (*Eichhornia crassipes*) that invaded Lake Victoria creates anoxic conditions restrictive to *L. niloticus* (Njiru et al., 2002) reducing population growth of *L. niloticus*. Meanwhile, water hyacinth provided refugia and favourable conditions for *C. gariepinus* to flourish (Barasa et al., 2014; Njiru et al., 2002). Therefore, reduced predation pressure and proliferation of *C. gariepinus* led to expansion and consequently stabilized the population. On the other hand, expansion of *C. gariepinus* population in Lakes Kamnarok and Kenyatta could be attributed to stock augmentation. In addition, the extensive coverage of Lake Kamnarok by *E. crassipes* reduced fishing pressure as well as provided ample feeding, breeding and nursery grounds for *C. gariepinus*, leading to population expansion.

The riverine *C. gariepinus* populations exhibited a uni-modal distribution suggesting a recent demographic expansion. This was consistent with the negative value of Tajima's D, indicative of a recent population size expansion after a bottleneck event. Indeed, the riverine *C. gariepinus* populations are prone to disturbances including decline in water levels due to drought, leaving individual fish in shallow pools, exposed to heavy fishing by human and predation by birds and anthropogenic activities such as sand harvesting (Masese et al., 2012), which destroy habitats for fish. However, recent episodes of flooding in the rivers (Masese et al., 2017) provided suitable conditions for the growth of the populations. Generally, flood conditions support proliferation of food items that support breeding and nursing of catfishes (Halwart & Gupta, 2004).

5 | CONCLUSIONS

Lakes Victoria, Kamnarok, Kenyatta and River Nyando *C. gariepinus* populations are reservoirs of higher genetic variation, while Rivers Tana and Sosiani *C. gariepinus* have lower genetic variation. All populations of *C. gariepinus* had singletons with Lakes Kenyatta, Kamnarok and River Nyando having higher number of singletons than Lake Victoria and Rivers Tana and Sosiani. Overall, these populations are important genetic resources, which should be conserved, by controlling factors likely to affect their genetic diversity such as overfishing, pollution, habitat destruction and fragmentation. Furthermore, more robust analysis by means of nuclear markers to supplement this work is recommended to validate these findings.

ACKNOWLEDGEMENTS

The National Commission for Science, Technology and Innovation (NACOSTI) (NACOSTI/RCD/ST&I 6th CALL MSc/078) and the National Research Fund (NRF) of Kenya (NRF/MSc/1/061) research grant awards to A.G.W funded the field sampling activities and part of the laboratory work for this study. The Austrian Partnership Programme in Higher Education and Research for Development

(APPEAR) for the STRECAFISH Project supported laboratory work at Makerere University's Molecular Biology Laboratory (MAKMBL) in Uganda by A.G.W. We thank Mr. J. Mayega of MAKMBL for technical support during laboratory work. Mr. G. Kasoso and Mr. J. Bett helped with field sampling in Lakes Kenyatta and Kamnarok, respectively.

CONFLICT OF INTEREST

Authors declare no conflict of interest in this manuscript.

DATA AVAILABILITY STATEMENT

Data for this work has been submitted to the GenBank database and is publicly available under the accession numbers MK014577 - MK014704.

ORCID

George W. Alal https://orcid.org/0000-0002-9788-1144
James E. Barasa https://orcid.org/0000-0001-7221-6618

REFERENCES

Barasa, J. E. (2018). Genetic diversity, population structure and influence on life-history traits of the African catfish, *Clarias gariepinus* in Kenya. PhD thesis, University of Eldoret, Kenya. 214 pages.

Barasa, J. E., Abila, R., Grobler, J. P., Agaba, M., Chemoiwa, E. J., & Kaunda-Arara, B. (2016). High genetic diversity and population differentiation in *Clarias gariepinus* of Yala Swamp: Evidence from mitochondrial DNA sequences. *Journal of Fish Biology*, *89*, 2557–2570. https://doi.org/10.1111/jfb.13150

Barasa, J. E., Abila, R., Grobler, J. P., Dangasuk, O. G., Njahira, M. N., & Kaunda-Arara, B. (2014). Genetic diversity and gene flow in *Clarias gariepinus* from Lakes Victoria and Kanyaboli, Kenya. *African Journal of Aquatic Science*, *39*, 287–293. https://doi.org/10.2989/16085914.2014.933734

Barasa, J. E., Mdyogolo, S., Abila, R., Grobler, J. P., Skilton, R. A., Bindeman, H., Ndotono, N. M., Chemoiwa, E. J., Dangasuk, O. G., Kaunda-Arara, B., & Verheyen, E. (2017). Genetic diversity and population structure of the African catfish, *Clarias gariepinus* (Burchell, 1822) in Kenya: Implication for conservation and aquaculture. *Belgian Journal of Zoology*, *147*, 105–127. https://doi.org/10.26496/bjz.2017.14

Barocca, T. M., Santos, G. B., Duarte, N. V. R., & Kalapothakis, E. (2012). Evaluation of genetic diversity and population structure in a commercially important freshwater fish *Prochilodus costatus* (Characiformes, Prochilodantidae) using complex hypervariable repeats. *Genetics and Molecular Research*, *11*, 4456–4467. https://doi.org/10.4238/2012.September.27.4

Bernatchez, L., & Danzmann, G. R. (1993). Congruence in control region sequence and restriction-site variation in mitochondrial DNA of Brook charr (*Salvelinus fontinalis* Mitchill). *Molecular Biology and Evolution*, *10*, 1002–1014. https://doi.org/10.1093/oxfordjournals.molbev.a040062

Bruton, M. N. (1979). The breeding biology and early development of *Clarias gariepinus* (Pisces, Clariidae) in Lake Sibaya, South Africa, with a review of breeding species of the subgenus Clarias. *Transactions of the Zoology Society of London*, *35*, 1–45. https://doi.org/10.1111/j.1096-3642.1979.tb00056.x

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

Chemoiwa, E. J., Abila, R., Macdonald, A., Lamb, J., Njenga, E., & Barasa, J. E. (2013). Genetic diversity and population structure of the endangered ripon barbell, *Barbus altianalis* in Lake Victoria catchment, Kenya, based on mitochondrial DNA sequences. *Journal of Applied Ichthyology*, *29*, 1225–1233. https://doi.org/10.1111/jai.12313

Chitamwebwa, D., Kamanyi, J., Kayungi, J., Nabbongo, H., Ogolla, A., & Ojuok, J. (2009). The present status of the hook fishery and its impact on fish stocks of Lake Victoria. *African Journal of Tropical Hydrobiology and Fisheries*, *12*, 78–82.

Corbet, P. S. (1961). The food of non-cichlid fishes in the Lake Victoria Basin, with remarks on their evolution and adaptation to lacustrine conditions. *Proceedings of the Zoological Society of London*, *136*, 1–101. https://doi.org/10.1111/j.1469-7998.1961.tb06080.x

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). JModelTest: Phylogenomics, Universidade de Vigo, Vigo, Spain (c) Grupo de Arquitectura de Computadores, UDC, A Coruna, Spain Project.

Excoffier, L. (2004). Analysis of population subdivision. *Handbook of Statistical Genetics*, https://doi.org/10.1002/0470022620.bbc25

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

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2002). *Introduction to Conservation Genetics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808999

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

Gathenya, M., Mwangi, H., Coe, R., & Sang, J. (2011). Climate and land use induced risks to watershed services in the Nyando River Basin. *Kenya. Exploratory Agriculture*, *47*(2), 339–356. https://doi.org/10.1017/S001447971100007X

Gene Codes Corporation. (2016). Sequencher 5 Series. www.genecodes.com

Grobler, J. P., Hoffman, L. C., & Prinsloo, J. F. (1997). A Comparison of Allozyme Heterozygosity and Life History variables in four strains of African Catfish (*Clarias gariepinus*). *Southern African Journal of Aquatic Sciences*, *23*(1), 31–41. https://doi.org/10.1080/10183469.1997.9631386

Hall, T. (2005). BioEdit v7. 0.5. Biological sequence alignment editor. Department of Microbiology, North Carolina State University. (Online) Available at: http://www.mbio.ncsu.edu/BioEdit/Bioedit.html

M. Halwart, & M. V. Gupta (Eds.) (2004). Culture of fish in rice fields. FAO and The WorldFish Center, 83 pages.

Kaufman, L., & Ochumba, P. (1993). Evolutionary and conservation biology of cichlid fishes as revealed by faunal remnants in northern Lake Victoria. *Conservation Biology*, *7*, 719–730. https://doi.org/10.1046/j.1523-1739.1993.07030719.x

Kayanda, R., Taabu, A. M., Tumwebaze, R., Muhoozi, L., Jembe, T., Mlaponi, E., & Nzungi, P. (2009). Status of the major commercial fish stocks and proposed species specific management plans for Lake Victoria. *African Journal of Tropical Hydrobiology and Fisheries*, *12*, 15–21. https://doi.org/10.4314/ajthf.v12i1.57366

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Lande, R. (1988). Genetics and demography in biological conservation. *Science*, *241*, 1455–1460. https://doi.org/10.1126/science.3420403

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X Version 2.0. *Bioinformatics*, *23*, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL v3): An online tool for the display and annotation of phylogenetic and oher relationships. *Nucleic acids Research*, *44*, 242–245.

Masese, F. O., Raburu, P. O., & Kwena, F. (2012). Threats to the Nyando Wetland. In *Community based Approach to the Management of Nyando Wetland* (pp. 68–80). Lake Victoria basin.

Masese, F. O., Salcedo-Borda, J. S., Gettel, G. M., Irvine, K., & McClain, M. E. (2017). Influence of catchment land use and seasonality on dissolved organic matter composition and ecosystem metabolism in headwater streams of a Kenyan river. *Biogeochemistry*, *132*(1–2), 1–22. https://doi.org/10.1007/s10533-016-0269-6

Muwanika, V., Nakamya, M., Rutaisire, J., Sivan, B., & Masembe, C. (2012). Low genetic differentiation among morphologically distinct *Labeobarbus* species (Teleostei: Cyprinidae) in the Lake Victoria and Albertine basins, Uganda: Insights from mitochondrial DNA. *African Journal of Aquatic Science*, *37*(2), 143–153. https://doi.org/10.2989/16085914.2012.668850

Ngugi, C. C., Omolo, B., Langdon, C., & Bowman, J. (2005). Studies on strategies for increasing the growth and survival of African catfish, Clarias gariepinus juveniles reared for stocking or for use as live bait. In: *25th Annual Technical report- Vol II*. Oregon State University, Snell Hall 400 Corvalis, Oregon 97331-1641, USA.

Njiru, M., Okeyo-Owuor, J. B., Muchiri, M., & Cowx, I. G. (2004). Shifts in the food of Nile tilapia, *Oreochromis niloticus* (L.) in Lake Victoria, Kenya. *African Journal of Ecology*, *42*, 163–170. https://doi.org/10.1111/j.1365-2028.2004.00503.x

Njiru, M., Othina, A. N., Getabu, A., Twiddle, D., & Cowx, I. G. (2002). Is the invasion of water hyacinth, *Eichhornia crassipes* solms (mart.), a blessing to Lake Victoria fisheries? In I. G. Cowx (Ed.), *Management and ecology of lake and reservoir fisheries* (pp. 255–263). Fishing News Books. https://doi.org/10.1002/9780470995679.ch21

Nyunja, C., Maina, J., Amimo, J., Kibegwa, F., Harper, D., & Jung'a, J. (2017). Stock structure delineation of the African Catfish (*Clarias gariepinus*) in selected populations in Kenya using mitochondrial DNA (Dloop) variability. *Journal of Aquaculture Research and Development*, *8*, 485. https://doi.org/10.4172/2155-9546.1000485

Ogutu-Ohwayo, R. (1990). The decline of the native fishes of lakes Victoria and Kyoga (East Africa) and the impact of introduced species, especially the Nile perch, *Lates niloticus*, and the *Nile tilapia*, *Oreochromis niloticus*. *Environmental Biology and Fisheries*, *27*, 81–96. https://doi.org/10.1007/BF00001938

Ojiambo, D. (2015). Population genetic structure of wild and domesticated African catfish (*Clarias gariepinus*) in Victoria and Albertine Drainage Basins. MSc. thesis, Makerere University, Kampala, Uganda. 53 pages.

Owiti, D. O., & Dadzie, S. (1989). Maturity, fecundity and the effect of reduced rainfall on the spawning rhythm of a siluroid catfish, *Clarias mossambicus* (Peters). *Aquaculture and Fisheries Management*, *20*, 355–368. https://doi.org/10.1111/j.1365-2109.1989.tb00363.x

Rogers, A. R., & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, *9*, 552–569.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DNAsP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*, 3299–3302. https://doi.org/10.1093/molbev/msx248

Sulem, Y. S., Tomedi, E. T., Mounchili, S., Tekeng, S., & Brummett, R. S. (2006). Survival of *Clarias gariepinus* fry in earthen ponds: Effects of composts and leaks. *Aquaculture*, *260*, 139–144. https://doi.org/10.1016/j.aquaculture.2006.06.008

Swofford, D. L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4. Sinauer Associates. Sunderland, Massachusetts.

Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics*, *123*, 597–601.

Van der Bank, F. H., Grobler, P. J., & Do Preez, H. H. (1992). A comparative biochemical genetic study of three populations of domesticated and wild African catfish (*Clarias gariepinus*). *Comparative Biochemistry and Physiology*, *101*(B), 387–390. https://doi.org/10.1016/0305-0491(92)90017-L

Vogl, A. L., Bryant, B. P., Hunink, J. E., Wolny, S., Apse, C., & Droogers, P. (2016). Valuing investments in sustainable land management in the Upper Tana River basin, Kenya. *Journal of Environmental Management*, *30*, 1–14. https://doi.org/10.1016/j.jenvman.2016.10.013

Witte, F., & van Densen, W. L. T. (1995). *Fish stocks and fisheries of Lake Victoria: A handbook for field observations*. Samara Publishing.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, *19*, 395–420. https://doi.org/10.1111/j.1558-5646.1965.tb01731.x

How to cite this article: Alal GW, Barasa JE, Chemoiwa EJ, Kaunda-Arara B, Akoll P, Masembe C. Genetic diversity and population structure of selected lacustrine and riverine populations of African catfish, *Clarias gariepinus* (Burchell, 1822), in Kenya. *J Appl Ichthyol*. 2021;00:1–12. https://doi.org/10.1111/jai.14167