Received: 23 August 2022 | Revised: 23 February 2023 | Accepted: 23 February 2023

DOI: 10.1111/aje.13144

RESEARCH ARTICLE

African Journal of Ecology WILEY

Genetic diversity of a cichlid fish population after 100 years of isolation

Titus Chemandwa Ndiwa[1] | Dorothy Wanja Nyingi[2] | Jean-François Agnèse[3]

[1]Department of Clinical Studies, University of Nairobi, Nairobi, Kenya

[2]Ichthyology Section, National Museums of Kenya, Nairobi, Kenya

[3]ISEM, University of Montpellier, CNRS, IRD, Montpellier, France

Correspondence
Titus Chemandwa Ndiwa, Department of Clinical Studies, University of Nairobi, P.O Box 29053-00625, Nairobi,
Email: tndiwa@uonbi.ac.ke

Funding information
IRD, France; National Museums of Kenya

Abstract

High rates of species radiation have been recorded in Cichlids of East African lakes. The current study focused on a population of Nile tilapia (*Oreochromis niloticus*) found in a crater lake (Lake Crocodile) that occurs within the Central Island of Lake Turkana. The population separated from Lake Turkana about 100 years ago and has since then existed in isolation. This study aimed at assessing the existence of genetic differences between the Lake Crocodile Nile tilapia population and other closely related Nile tilapia populations from Lake Turkana, Lake Baringo and River Suguta. Genetic differences between the populations were evaluated using two mtDNA markers, and 16 microsatellite markers. Significant genetic differences ($p > 0.05$) were observed between the Lake Crocodile population and the other studied populations within the region. Despite the shared ancestry between Lake Crocodile and Lake Turkana, genetic differences were observed between the two populations. This can mostly be attributed to the lack of gene flow and environmental differences between the two populations.

KEYWORDS
cichlid, genetic differentiation, Island, Lake Crocodile, Nile tilapia

Abstract

Des taux élevés de radiation des espèces ont été enregistrés chez les Cichlidés des lacs d'Afrique de l'Est. L'étude actuelle s'est concentrée sur une population de tilapia du Nil (Oreochromis niloticus) trouvée dans un lac de cratère (Lac Crocodile) situé dans l'île centrale du Lac Turkana. La population s'est séparée du Lac Turkana il y a environ 100 ans et vit depuis lors dans l'isolement. Cette étude visait à évaluer l'existence de différences génétiques entre la population de tilapia du Nil du Lac Crocodile et d'autres populations de tilapia étroitement apparentées du Nil du Lac Turkana, du Lac Baringo et du Fleuve Suguta. Les différences génétiques entre les populations ont été évaluées à l'aide de deux marqueurs ADNmt et de 16 marqueurs microsatellites. Des différences génétiques considérables ($p > 0.05$) ont été observées entre la population du Lac Crocodiles et les autres populations étudiées dans la région. Malgré l'ascendance commune entre le Lac Crocodile et le Lac Turkana, des différences génétiques ont été observées entre les deux populations. Cela peut être attribué principalement à l'absence de flux génétique et aux différences environnementales entre les deux populations.

1 | INTRODUCTION

Lake Turkana, previously known as Lake Rudolf is the largest lake on the Eastern arm of the Rift Valley. Compared with the three largest African lakes (Victoria, Tanganyika and Malawi), characterised by high number of endemic cichlids, Lake Turkana only has seven endemic species (Butzer, 1971). The fish community within the lake is mainly made up of Nilotic riverine fish fauna, and characterised by low species diversity (Kolding, 1989). According to a study by Kolding (1993), only three cichlid species are known to inhabit the lake; *Sarotherodon galilaeus* (Mango tilapia), *Coptodon zillii* and *Oreochromis niloticus* (Nile tilapia).

A research study based on morphometric and meristic characters by Trewavas (1983) described seven different subspecies of *O. niloticus*. The study reported that the *O. niloticus* subspecies associated with the Lake Turkana basin is *O. niloticus vulcani*, while the *O. niloticus* populations from Lake Baringo and River Suguta belonged to *O. niloticus baringoensis* and *O. niloticus sugutae* subspecies, respectively. Later studies by Nyingi et al. (2009) demonstrated that the three *O. niloticus* subspecies were genetically differentiated.

During the study by Trewavas (1983), the *O. niloticus vulcani* subspecies in Lake Turkana consisted of three subpopulations that included fishes within the main lake (Lake Turkana population), and two Crater lake populations in Central Island within Lake Turkana. The island had three small Crater lakes, with two of these lakes supporting populations of *O. niloticus*. According to Trewavas (1983), these two "in lake" isolated populations shared the same basic morphometric and meristic characteristics with *O. niloticus vulcani* subspecies from Lake Turkana. Consequently, Trewavas (1983) assigned these two populations to *O. niloticus vulcani* subspecies. However, recent studies by Ndiwa (2014) and Ndiwa et al. (2016) comparing the morphological characteristics of the Lake Crocodile populations to Lake Turkana population showed that they had significant morphological differences. The observed differences were attributed to variations in environmental characteristics within the two habitats (Ndiwa, 2014; Ndiwa et al., 2016).

Studies within the two study sites revealed that water levels in Lake Turkana vary constantly depending on the amount of precipitations and temperatures experienced in the region. Lake Turkana has experienced significant episodes of climatic change since late Quaternary and Holocene (Ambrose, 2001; Garcin et al., 2009; Mworia, 1991; Umer et al., 2004). The latest episode is associated with the high decline in water level and is believed to have occurred around 1900. The drastic decline of water level in Lake Turkana exposed and isolated the three Crater Lakes within the Central Island, which have remained separated from the main lake to date (Hopson, 1982). The fishes that remained trapped within the newly formed crater lakes (Lake Tilapia and Lake Crocodile) formed new populations of *O. niloticus vulcani*. However, reports from recent surveys within the Central Island have indicated that for unknown reasons, the Nile tilapia population within Lake Tilapia became locally extinct, while the Lake Crocodile population continues to flourish.

Lakes Crocodile and Turkana though located within the same drainage system differ in many aspects (Ndiwa, 2014; Ndiwa et al., 2016). Lake Turkana is one of the oldest and largest closed basins covering an area of approximately 7500 km^2, and has an average depth of 35 m (Owen et al., 1982; Yuretich, 1979). Its waters are moderately saline (2500 mg/L) and alkaline having a pH of about 9.2. Approximately 80% of the water in the lake comes from Ethiopian Highlands through Omo River (Yuretich & Cerling, 1983). Other perennial rivers that feed the lake include Rivers Turkwel, Kerio, Ileret and Tula-Bor (Frostick & Reid, 1990). On the contrary, Lake Crocodile being a crater lake within an Island in Lake Turkana, has an area of <1 km^2. Water from Lake Turkana percolates into the crater lake. Therefore, the amount of water in Lake Crocodile depends on the water level in Lake Turkana. Salinity in Lake Crocodile is about three

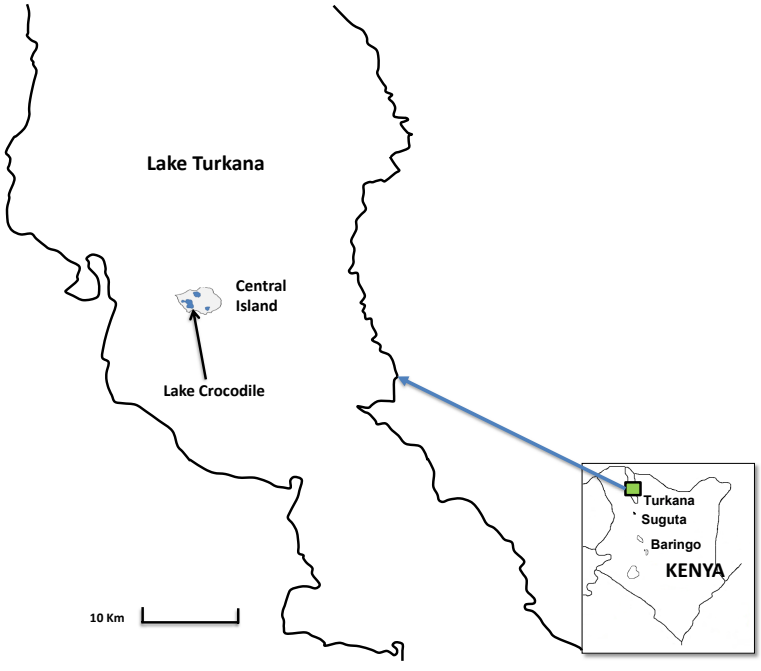

FIGURE 1 Map showing the location of the study area (Lake Turkana and Lake Crocodile), and relative location of the other two Nile tilapia populations (Suguta and Baringo).

times more than salinity in Lake Turkana. Similarly, the crater lake has been reported to have higher water pH of about 10 compared to Lake Turkana (Avery, 2010). The current study aimed at assessing genetic impact of separation and new environmental conditions on the population of *O. niloticus* in Lake Crocodile. The two main questions investigated were (i) Are there significant genetic differences between Lake Crocodile Nile tilapia population and other closely related Nile tilapia populations? (ii) How can the observed genetic differences be explained? To answer these important questions, partial mtDNA sequences and 16 microsatellite loci were studied.

2 | MATERIALS AND METHODS

The Lake Turkana system is located in the Northern part of Kenya at 3′ N, 36° E. Crocodile Lake on the other hand is a Crater Lake located on Central Island within Lake Turkana (Figure 1). The two other populations (Baringo and Suguta) are located on the southern part of Lake Turkana, at coordinates 0°36′ N, 36°01′ E and 1° 10′ N, 36°06′ E, respectively (Figure 1).

2.1 | Collection of samples

A total of 111 fish samples were collected from Lakes Baringo (30), Turkana (30), Crocodile (30) and River Suguta (21) using gill nets and seine nets after obtaining permission from Kenya Wildlife Service (permit number KWS/BRM/5001). Nile tilapia specimens were identified based on morphological characteristics, and immediately sacrificed using an overdose of M-22, in order to minimise the suffering of the fish. DNA tissue samples were collected from both the fins and muscles and were preserved in 95% ethanol. The fish specimens were fixed using 4% formaldehyde solution and curated at Ichthyology section of the National Museums of Kenya (Voucher No. FW/4485/1-40).

2.2 | DNA extraction and sequencing

Genomic DNA extraction was carried out using the salting out method as described by Aljanabi and Martinez (1997). The extracted genomic DNA was amplified using mtDNA markers (Table 1) on a final volume of 50 μL comprising 0.25 mM $MgCl_2$, 0.2 mM of each dNTP, 1 mM of each primer, 5 mL of 10X buffer and 10 units of Taq polymerase (Promega). The amplification reaction conditions entailed pre-heating for 3 min at 93°C followed by 40 cycles of 30 s at 93°C, 30 s at 62°C and 1 min at 72°C, with a final elongation phase of 72°C for 5 min. The resulting products were sequenced on an ABI 3730 XL automated sequencer at the Centre Méditerraéen de l'Environnement et de la Biodiversite (CeMEB) laboratory in Montpellier, France.

The extracted genomic DNA was amplified at 16 microsatellite loci (Table 1). The used microsatellite markers were developed and used in previous studies by Agnese et al. (2008), Carleton et al. (2002) and Lee and Kocher (1996). The QIAGEN multiplex PCR kit (www.qiagen.com) and fluorescently labelled primers were used to amplify the microsatellite loci following the manufacturer's protocol. The markers were pooled into four groups as follows: group 1 (UNH1003, UNH115, UNH154 and UNH860), group 2 (UNH 211, UNH104, UNH146 and UNH995), group 3 (PRL1GT, UNH 142, UNH189 and UNH887) and group 4 (PRL1AC, UNH129, UNH162 and UNH874). The forward primers in each group were then labelled using fluorescent dyes NED, PET, VIC or FAM. The PCR reaction conditions involved 5 min pre-heating at 95°C, followed by 30 cycles of 30 s at 95°C, 90 s at 60°C (except group 1 markers; 50°C) and 1 min at 72°C, with a final elongation phase of 5 min at 72°C. 3 mL of PCR products were diluted in 12 mL of HiDi formamide and 0.2 mL of 500 LIZ size standard added. Finally, electrophoresis was carried out in an ABI 3730 XL automated sequencer and the resulting electrophorograms analysed using GeneMapper to determine allele sizes. The microsatellite loci for all the lakes were generated under same study and were scored by same individual in order to minimise scoring errors.

The mtDNA sequences belonging to the other two closely related neighbouring Nile tilapia populations from Lake Baringo and River Suguta were obtained from Genbank and were used for comparison purposes.

2.3 | Data analysis

The new sequences and those obtained from Genbank were aligned using SEAVIEW version 4.3.5 (Gouy et al., 2010). Genetic diversity index was then calculated using DnaSP version 5.10.01 (Librado & Rozas, 2009). Based on DnaSP, genetic diversity was analysed by computing average number of nucleotide differences per site between two sequences (π), and the average number of nucleotide differences between sequences (k). The neutrality index of Tajima's D (Tajima, 1989) was also calculated for each Nile tilapia population.

The relationships among the studied populations were evaluated and depicted on a TSC haplotype network using PopART version 1.7 (Clement et al., 2000; Leigh & Bryant, 2015).

All microsatellite loci were checked using MICRO-CHECKER version 2.2.3 (Van Oosterhout et al., 2004) in order to investigate the presence of null alleles, and scoring errors that may arise due to stuttering or large alleles.

Factorial Component Analysis (FCA) was carried out using GENETIX version 4.05 (Belkhir et al., 2004) to establish maximum genetic variations between studied populations, and visualise these differences.

Basic data on genetic composition (alleles per locus, expected heterozygosity and observed heterozygosity) of the populations was computed using GENETIX version 4.05 (Belkhir et al., 2004). F-statistics and tests for Hardy–Weinberg were calculated using FSTAT version 2.9.3.2 (Goudet, 1995a, 1995b) at 95% confidence interval.

TABLE 1 mtDNA and 16 microsatellite markers used in the study.

Marker		Primer (forward/reverse)
mtDNA (d-loop)		F 5′-ACCCCTAGCTCCCAAAGCTA-3′
		R 5′-CCTGAAGTAGGACCAGATG-3′
Microsatellite	**Genbank number**	**Primer sequence (F and R)**
PRL1AC		F 5′-TCGTGTCTTGTGGGGAAACC-3′
		R 5′-TGAATGGA TGCAACAGGATG-3′
PRL1GT		F 5′-GTTAGCCCCCTCCTCACTCT-3′
		R 5′-ACCTTGCTCGTCACACCTG-3′
UNH 104	G12257.1	F 5′-GCAGTTATTTGTGGTCACTA-3′
		R 5′-GGTATATGTCTAACTGAAATCC-3′
UNH 115	G12268.1	F 5′-ACCTTCATCTCGGTCAG-3′
		R 5′-TCAAGCAGCTGATTTTT-3′
UNH 129	G12282.1	F 5′-AGAAGTCGTGCATCTCTC-3′
		R 5′-TGTACATCATCTGTGGG-3′
UNH 142	G12294.1	F 5′-CTTTACGTTGACGCAGT-3′
		R 5′-GTGACATGCAGCAGATA-3′
UNH 146	G12298.1	F 5′-CCACTCTGCCTGCCCTCTAT-3′
		R 5′-AGCTGCGTCAAACTCTCAAAAG-3′
UNH 154	G12306.1	F 5′-ACGGAAACAGAAGTTACTT-3′
		R 5′-TTCCTACTTGTCCACCT-3′
UNH 162	G12314.1	F 5′-CAGACACAGCAGAGGAT-3′
		R 5′-TGATAAGTAATTCATCTGTTTG-3′
UNH 189	G12341.1	F 5′-ATCGATGCTTTAAGAATCAG-3′
		R 5′-TTCTCTGACATTTTTCAGC-3′
UNH 211	G12362.1	F 5′-GGGAGGTGCTAGTCATA-3′
		R 5′-CAAGGAAAACAATGGTGATA-3′
UNH 860	G68195.1	F 5′-ACTGTTTACCCACTGCGACA-3′
		R 5′-AGATGTGTCTGAGCCATCCA-3′
UNH 874	G68202.1	F 5′-AGTAAAATGGGCGAACGTGT-3′
		R 5′-TGAAGCTGGGAGTTTCCTGT-3′
UNH 887	G68210.1	F 5′-ACCTTTCGTACAGGGCACAG-3′
		R 5′-CCACAATGACAACTCAACACC-3′
UNH 995	G68274.1	F 5′-CCAGCCCTCTGCATAAAGAC-3′
		R 5′-GCAGCACAACCACAGTGCTA-3′
UNH 1003	G68280.1	F 5′-CAGTGTTAAGTGGCTTCACCA-3′
		R 5′-AGCAAGGAACTCGAGAGCTG-3′

Abbreviations: F, forward primer; R, reverse primer.

To test for genetic bottleneck signature in the four populations, the program BOTTLENECK version 1.2.02 (Cornuet & Luikart, 1996; Piry et al., 1999) was used. This analysis was based on three models of mutation; Two-Phase Model (TPM), Infinite Allele Model (IAM) and Single Step Model (SSM). For the TPM, 90% of mutation was defined to follow the stepwise mutation model, and 10% as multistep variance (Garza & Williamson, 2001; Hundertmark & Van Daele, 2010). Excess in heterozygosity indicates presence of bottleneck in a population.

3 | RESULTS

3.1 | mtDNA results

The current study generated 16 new sequences for the Lake Crocodile population and compared them to 38 sequences obtained from genbank (Table 2).

Analysis of these sequences showed that all populations studied were polymorphic. A total of 23 different haplotypes were

TABLE 2 Genbank numbers of the analysed sequences.

Lake	GenBank number	Voucher number
Turkana	EF016680-016696, KJ746045-746050	(NMK) 1639/1-11,1641/1-13, (ISEM) JFA-2010-LT1-30
Crocodile	MG018341 – MG018356	(NMK) FW/4485/1-40
Suguta	EF016709-016714	(ISEM) JFA-5287-5313
Baringo	EF016697-016708, FJ440604-440607, KJ746041-44	(NMK) FW/2660/1-31, MRAC 95-027-P-0074-0084

TABLE 3 Genetic diversity and neutrality index (Tajima's D) of the four populations of Nile tilapia based on mtDNA sequences analysis (number of sequences; *n*, number of haplotypes; H_{ob}, number of polymorphic sites; *p*, haplotype diversity; H_d, average number of nucleotide differences; *k*, and nucleotide diversity; π).

Parameter/population	Suguta	Baringo	Turkana	Crocodile
Number of sequence (*n*)	6	16	16	16
Number of haplotype (H_{ob})	4 (1)	7 (5)	15 (11)	4 (1)
Number of polymorphic sites (*p*)	7	241	15	4
Haplotype diversity (H_d)	0.80000	0.95000	0.90833	0.79167
Average number of nucleotide difference (*k*)	2.33333	94.2500	2.27500	1.84167
Nucleotide diversity (π)	0.00697	0.28134	0.00679	0.00550
Tajima's D	−1.36732	0.63863	−1.82162*	1.65113

Note: Number of unique haplotypes are enclosed in brackets, and statistical significance is denoted by an asterisk (*).

observed, with highest number of haplotypes (15) occurring in Lake Turkana (Table 3). Lake Baringo had seven haplotypes, while the two other populations (Crocodile and Suguta) had the lowest number of haplotypes (four and three, respectively).

From these haplotypes, 18 were unique (Figure 2; Table 3). Lakes Turkana and Baringo populations recorded the highest number of unique haplotypes (11 and 5, respectively), while Lakes Crocodile and Suguta had the lowest number of unique haplotypes (each population had one unique haplotype) (Figure 2; Table 3).

Analysis of the 54 mtDNA sequences using the TCS Network (Figure 2) indicated that the Lake Crocodile population was closely related to Lake Turkana and Lake Baringo populations. There were two peripherally located haplotypes from Lake Crocodile Nile tilapia population, indicating their recent formation.

The Lake Crocodile population showed the lowest genetic diversity compared to the other three populations studied (Table 3). Its nucleotide and haplotype diversities were 0.00550 and 0.79167, respectively. The other three populations had slightly higher nucleotide and haplotype diversities ranging from 0.00679–0.28134 and 0.80000–0.95000, respectively. The highest nucleotide diversity was observed in the Lake Baringo population (Table 3).

Neutrality index calculated by Tajima's D indicated negative values for R. Suguta and Lake Turkana populations, with the latter being significant (Table 3). Lakes Crocodile and Baringo populations had positive D values, which were not significant (Table 3).

3.2 | Microsatellite results

Analysis of microsatellite data on Micro-checker found evidence of null alleles at locus UNH 115 for one population, and a few individuals with missing data. The locus and the individuals with missing data were therefore excluded from downstream analysis. Overall, the most diverse locus was UNH 154, while the least diverse locus was UNH 129. Further analysis indicated that Lake Turkana had the highest allelic richness (*A*), followed by Lake Crocodile and Lake Baringo populations which had almost equal mean allelic richness (Table 4). However, River Suguta population had the least mean allelic richness (*A*), observed heterozygosity (H_o) and expected heterozygosity (H_e) (Table 4).

The pairwise *Fst* values of studied populations were all significant ($p < 0.05$), and ranged from 0.0236–0.1742 (Table 5). As expected, the lowest pairwise *Fst* values were observed between Lakes Crocodile and Turkana population, an observation attributed to their recent isolation. Alternatively, the highest level of genetic differentiation was observed between Lakes Crocodile and Suguta populations (Table 5). Suguta population had similar *Fst* value in relationship to Lake Baringo and Lake Turkana populations.

Genetic diversity differed between the 16 loci and populations as shown in Table 6. Mean genetic diversity was highest at Lake Baringo (0.838 ± 0.0192), followed by Lake Turkana with mean genetic diversity of 0.824 ± 0.0330. Crocodile Lake population had low-genetic diversity (0.796 ± 0.0270), while Lake Suguta population recorded the least mean genetic diversity of 0.684 ± 0.0429. Additionally, the least diverse locus was UNH 860 (0.191) in Suguta population, while the most diverse locus was UNH 154 (0.961) in Turkana population. Similarly, UNH 154 was the most diverse locus among the studied populations except Baringo where UNH 211 had the highest diversity (0.948). The least diverse loci were UNH 860 (Suguta, 0.191), UNH 1003 (Crocodile, 0.585), and UNH 129 in Turkana (0.553) and Baringo (0.627) populations.

Factorial component analysis of the genotypes of the four populations studied (Figure 3) showed distinct genetic differentiation

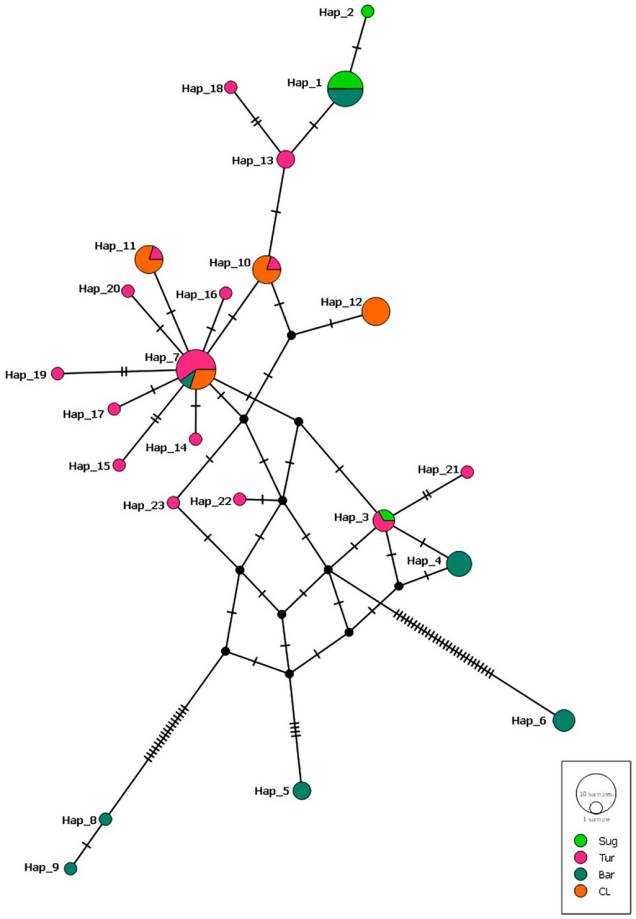

FIGURE 2 TCS haplotype network of the mtDNA sequences of Nile tilapia from Lake Crocodile (CL), Lake Turkana (Tur), Lake Baringo (Bar) and River Suguta (Sug). The size of the circle is proportional to the haplotype frequency.

TABLE 4 Genetic diversity of the microsatellites of samples of four populations of Nile tilapia studied (*A*, allelic richness; *n*, sample size; H_o, observed heterozygosity; H_e, expected heterozygosity, F_{is}, inbreeding coefficient).

Poulation	*n*	*A*	H_o	H_e	F_{is}
Crocodile	30	11.3125	0.7509	0.7816	0.0556
Turkana	30	13.5625	0.7607	0.8080	0.0765
Baringo	30	11.2500	0.7935	0.8229	0.0536
Suguta	21	5.8125	0.6289	0.6643	0.0628

between all the studied populations. The differentiation of the populations was mainly based on axis 1, which accounted for 49.28% of the observed variations. The Baringo and Suguta populations were discriminated by second axis (32.71%) and third axis (18.01%), while Turkana–Crocodile populations were mainly discriminated by the latter axis. Lake Crocodile population was more similar to Turkana than Baringo and Suguta. Further analysis of genotypes from Crocodile and Turkana populations (Figure 4) indicated clear genetic separation of the two populations based on first axis (100%).

TABLE 5 *Fst* values calculated using FSTAT test of the four populations studied.

Population	Crocodile	Turkana	Baringo
Turkana	0.0234*		
Baringo	0.1276*	0.1096*	
Suguta	0.1716*	0.1368*	0.1368*

Note: Asterix * indicative significant difference at 95% confidence interval.

TABLE 6 Results of gene diversity per locus in each of the studied populations (The least and most diverse locus in each population are highlighted in bold).

Loci/Population	Crocodile	Turkana	Baringo	Suguta
PRL1AC	0.805	0.921	0.902	0.815
PRL1GT	0.724	0.864	0.893	0.729
UNH 211	0.889	0.879	**0.948**	0.747
UNH1003	**0.585**	0.721	0.837	0.767
UNH104	0.882	0.911	0.769	0.795
UNH115	0.667	0.667	0.804	0.771
UNH129	0.638	**0.553**	**0.627**	0.595
UNH142	0.865	0.806	0.817	0.721
UNH146	0.703	0.568	0.892	0.617
UNH154	**0.939**	**0.961**	0.920	**0.907**
UNH162	0.844	0.877	0.889	0.660
UNH189	0.911	0.916	0.855	0.437
UNH860	0.867	0.946	0.816	**0.191**
UNH874	0.813	0.917	0.849	0.785
UNH887	0.720	0.764	0.827	0.620
UNH995	0.882	0.914	0.765	0.784
Mean	0.7959	0.8241	0.8381	0.6838

3.3 | Bottleneck tests

Further analysis was carried out to establish occurrence of bottlenecks in the studied populations (Table 7). Lake Crocodile population showed significant cases of excess heterozygosity using sign test (TPM and SMM), standards different test (IAM, TPM and SMM) and Wilcoxon test (SMM). Similarly, significant excess heterozygosity was also observed using sign test (SMM), standard difference test (TPM and SMM) and Wilcoxon test (IAM and SMM) on Lake Turkana population (Table 7). Lake Baringo and River Suguta populations showed significant excess heterozygosity based on IAM in all the three tests (Table 7).

4 | DISCUSSION

4.1 | Genetic characterisation of the Nile tilapia populations

Genetic characterisation studies based on TCS network indicated that the Lake Crocodile population was closely related to the Lake

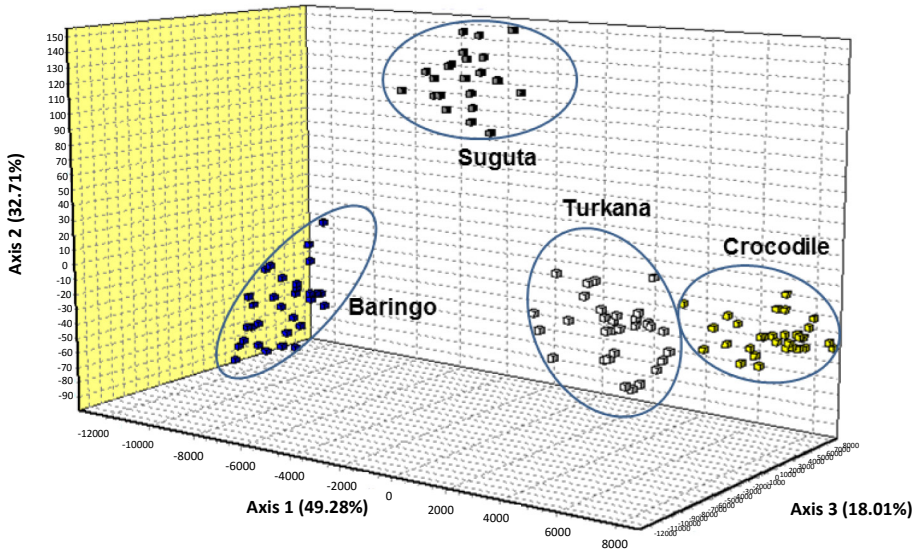

FIGURE 3 Results of factorial component analysis showing the relationship between the four populations of Nile tilapia studied.

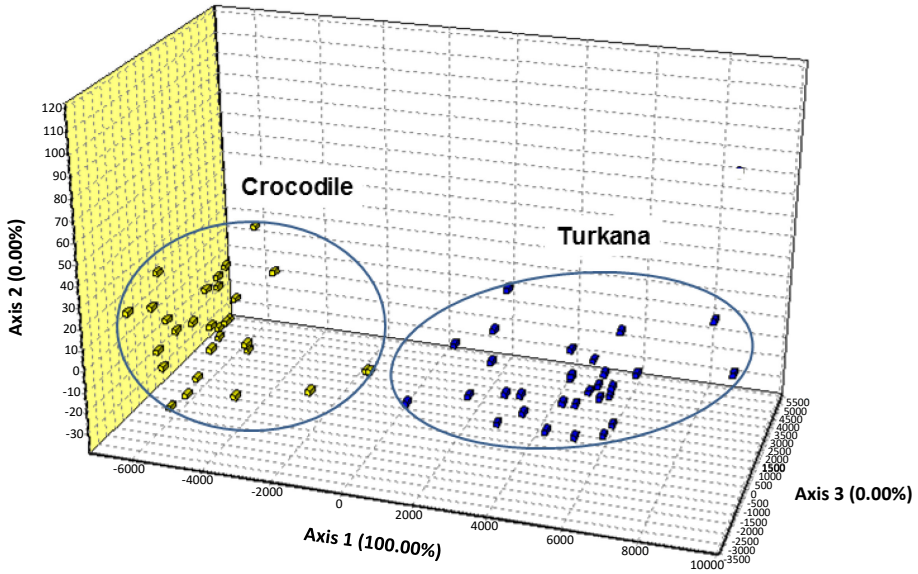

FIGURE 4 Factorial component analysis results showing distinct separation of Crocodile population from Turkana population.

Turkana and Lake Baringo populations. However, there were no haplotypes shared between Lake Crocodile and River Suguta Nile tilapia populations, an indication of their distant relationship. Indeed, Lake Baringo and River Suguta populations have been reported to have existed in isolation from the Lake Turkana population for more than 121,000 years (Coe, 1971; Fuchs, 1950).

Even though genetic diversity analysis indicated occurrence of lowest diversity at Lake Crocodile Nile tilapia population (Table 3), its mean allelic richness was higher than the mean allelic richness observed in the Nile tilapia population from River Suguta (Table 4). According to Coe (1971) and Fuchs (1950), River Suguta was part of the Lake Turkana basin, but was isolated during Middle and Upper Pleistocene volcanic activities. The low-diversity recorded in the current study may be attributed to the environmental conditions of the River, and its seasonal nature. The river is characterised by diverse environmental conditions. It receives water from hot-springs (too hot for fish to survive), which becomes warm and alkaline as the water flows away from the springs. Fishes inhabit areas of the river with temperatures ranging between 32°C and 38°C (Coe, 1971). The river is also seasonal; it flows during the rainy season, and its water reduces drastically as it dries up during the dry season. These occurrences predispose the population to constant episodes of bottleneck cycles resulting to low-genetic diversity.

Compared with the other three populations of Nile tilapia, Lake Baringo population had relatively higher genetic diversity (Tables 3 and 4). The unusual high-genetic diversity in the lake had earlier been observed by Nyingi et al. (2009). According to the authors, the high-genetic diversity was as a result of genetic introgression from *Oreochromis leucostictus* species introduced into Lake Baringo from Lake Naivasha.

TABLE 7 Results of bottleneck analysis based on three mutational models; IAM, TPM and SMM.

Population	Sign test			Standard difference test			Wilcoxon test		
	IAM	TPM	SMM	IAM	TPM	SMM	IAM	TPM	SMM
Crocodile	0.1786	**0.0185**	**0.0055**	**0.0385**	**0.0000**	**0.0000**	0.0507	0.0577	**0.0021**
Turkana	0.0694	0.1280	**0.0196**	0.0891	**0.0003**	**0.0000**	**0.0335**	0.1167	**0.0250**
Baringo	**0.0004**	0.5000	0.0623	**0.0002**	0.4459	0.1179	**0.0000**	0.8603	0.1754
Suguta	**0.0179**	0.5060	0.3174	**0.0015**	0.4752	0.1238	**0.0001**	0.8999	0.5966

Note: In bold are tests that showed significant excess heterozygosity (bottleneck events).
Abbreviations: IAM, Infinite Allele Model; SMM, Stepwise Mutation Model; TPM, Two-Phase Model.

4.2 | Loss of genetic variability after isolation of Crocodile population

A lower genetic variability was observed in the Lake Crocodile population compared to Lake Turkana population (Table 3). Since the Lake Crocodile founder population is thought to have come from Lake Turkana, the observed decline in genetic diversity as measured by the number of alleles and heterozygosity may be attributed to the small size of the Nile tilapia population in Lake Crocodile, genetic drift and restricted gene flow. This observation is supported by occurrence of population bottlenecks (Table 7) in Crocodile Lake.

Normally, recently isolated populations such as Lake Crocodile are always small in size. The small surface area of the Lake Crocodile (<1 km^2) further limits the number of individuals of Nile tilapia that can inhabit the lake. The low-genetic variability observed in the current study supports the findings of Frankham (1996), who noted that a positive correlation exists between genetic variability and population size within wildlife species. Generally, lakes with large surface areas (Lake Turkana and Lake Baringo) had higher genetic diversities, compared to Lake Crocodile and River Suguta which had smaller sizes.

The decline in genetic variability at Lake Crocodile population may also be attributed to selection pressure on founding individuals under the new environment after colonisation and subsequently isolation of the new population. Indeed, water in Lake Crocodile is characterised by high water pH (10) and salinity having conductivity of about 10,590 μS/cm, compared to Lake Turkana with both lower water pH (9.3) and salinity having conductivity of about 3420 μS/cm (Avery, 2010; Ndiwa, 2014; Ndiwa et al., 2016). The high pH and salinity under the new environment might have acted as selecting pressures resulting to loss of some genes, especially those not suitable for survival under the new environment hence reducing genetic variability. Such selection pressures that exist in new environments favour divergence of populations from their ancestral forms into new forms. The current study also noted occurrence of a unique haplotype within the Lake Crocodile Nile tilapia population. The presence of this unique haplotype may be attributed to isolation and subsequent diversification that may have resulted from above discussed selection pressures, or lack of sampling of the haplotype within Lake Turkana population.

Small newly formed populations, such as the Lake Crocodile Nile tilapia population play a significant role in conservation. According to studies by Ledoux et al. (2021), small newly formed populations are considered as potential reservoirs for adaptive genetic variation despite their low-genetic diversity. In the current global warming scenario, it is important for organisms to adapt to the changing environment for their survival. Additionally, adaptations of the Lake Crocodile Nile tilapia to high salinity and pH makes this population an important aquacultural resource. This can be achieved by using them as natural models for understanding adaptation of Nile tilapia to high salinity and pH, and also exploiting their genetic resources in development of tilapia aquaculture.

4.3 | Genetic differentiation of Crocodilepopulation

Factorial component analysis analysis separated the four populations of Nile tilapia into four clusters (Figure 3). This isolation was mainly based on genetic isolation by distance, and time. Lake Crocodile is closely related to Lake Turkana due to their recent isolation of about 100 years (Hopson, 1982) compared to Lake Baringo and River Suguta, which are thought to have separated from Lake Turkana during Middle and Upper Pleistocene volcanic activities (Coe, 1971; Fuchs, 1950).

Further genetic analysis based on FCA clustered Lake Turkana and Lake Crocodile populations into two distinct groups. Clear genetic differentiation between these two recently isolated populations can be attributed to lack of gene flow between the two populations for the past 100 years. In evolution, 100 years may be relatively short time for such significant differentiation to occur. However, various studies have also reported cases of rapid diversification of cichlids. In East African Lakes, for instance, past studies by different authors (Meyer et al., 1990; Owen et al., 1990; Sage et al., 1984) reported that cichlid species went through a period of rapid morphological diversification and speciation, leading to formation of many new species within a few thousand generations. The current study, therefore corroborates these findings. The observed genetic differentiation between the studied populations may also be attributed to bottleneck events. The current study noted that the Nile tilapia populations had experienced population bottlenecks. The bottleneck events can contribute to drastic changes in genetic composition of populations resulting to highly differentiated populations as observed in our study.

Neutrality test based on Tajima's D recorded negative D values for Lake Turkana and River Suguta populations. This observation indicate that the two populations might have experienced recent expansions, or undergone genetic selection that reduces variation. On the contrary, Lake Crocodile and Lake Baringo populations recorded positive D values implying that these populations might have undergone recent contraction, or experienced selection that enhanced their genetic variation.

This study employed the use of microsatellites markers in evaluating genetic diversity and genetic differentiation between the Nile tilapia populations. Microsatellite markers have widely been used in the study of population genetics, conservation genetics and evolutionary studies (Barba et al., 2016; Gonzalez-Castellano et al., 2020). Microsatellite markers are codominant in nature, have biparental mode of inheritance, and have high levels of polymorphism due to their multi-allelic nature (Allendorf, 2017). As a result, various studies have been conducted using microsatellites to assess population structure (Heras et al., 2019; Jorde et al., 2015; Perina et al., 2019). Over the recent past, use of single nucleotide polymorphisms (SNPs) in population genetics has also developed as an extensively used genomic tool of assessing patterns of genetic variations in populations. This is attributed to the higher number of analysed loci that results in a more extensive screening of genome, a lower genotyping error, a higher reproducibility among laboratories and a more precise estimate of diversity attributed to the high rate of mutation of microsatellites (Helyar et al., 2011). However, the use of SNPs requires a large number of markers for attainment of robust estimates of genetic variations among populations. The use of microsatellites on the other hand does not require large number of markers as a result of their multi-allelic nature that provide higher levels of polymorphism, high-evolutionary rate and low levels of ascertainment bias (Allendorf, 2017). Further studies can be conducted on the studied Nile tilapia populations using SNP markers for comparison purpose. Additionally, SNPs can be used to evaluate if there are variants specific to Lake Crocodile population that might be related to adaptation to its extreme environmental conditions such as pH and salinity.

5 | CONCLUSION

Previous morphological studies on various populations of Nile tilapia living in habitats characterised by extreme environmental conditions recorded clear morphological differences between Lake Crocodile and Lake Turkana populations (Ndiwa, 2014; Ndiwa et al., 2016). These morphological differences were partly suspected to have been influenced by genetic factors. The findings of the current study therefore confirm congruence in the observed morphological and genetic differences. These provides evidence that the Lake Crocodile Nile tilapia population is on independent evolutionary trajectory hence the need for separate management and conservation from the Lake Turkana Nile tilapia population.

AUTHOR CONTRIBUTIONS

All the authors contributed equally to this work.

ACKNOWLEDGEMENTS

We sincerely thank Fre'de'rique Cerqueira, Erick Desmarais (ISEM), Joseph Gathua (NMK), Dr. Felix Kibegwa and Dr. Judith Nyunja (KWS) for their technical support that ensured successful completion of this study. We are also grateful to the following organisations; Centre Méditerraéen de l'Environnement et de la Biodiversité (CeMEB), Institut des Science de l'Evolution de Montpellier (ISEM), National Museums of Kenya (NMK) for allowing us use their facilities, and KENWEB, Institut Recherche pour le Développement (IRD) and French Embassy in Nairobi for their financial support. We also thank the editors and the anonymous reviewers who took time out of their busy schedules to go through our manuscript, and provide insightful review comments that greatly enhanced the quality of this article.

CONFLICT OF INTEREST STATEMENT

The authors of this article have no conflict of interest.

DATA AVAILABILITY STATEMENT

Fish specimens from which tissues were obtained or tissue samples are curated at National Museums of Kenya voucher number FW/4485/1-40, NMK 1639/1-11, NMK 1641/1-13, NMK 2660/1-31, and ISEM laboratory under voucher numbers JFA-2010-LT1-30, JFA-5287-5313 and MRAC 95-027-P-0074-0084. Genetic sequences used in the study are available on genbank under accession numbers EF016680-016696, KJ746045-746050, MG018341 – MG018356, EF016709-016714, EF016697-016708, FJ440604-440607 and KJ746041-44.

ORCID

Titus Chemandwa Ndiwa https://orcid.org/0000-0001-5262-6142

REFERENCES

Agnese, J. F., Adepo-Gourene, B., & Nyingi, D. (2008). Functional microsatellite and possible selective sweep in the natural populations of the black-chinned tilapia *Sarotherodon melanotheron* (Teleostei, Cichlidae). *Marine Genomics*, *1*(3–4), 103–107.

Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, *25*(22), 4692–4693.

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: Allozymes to genomes. *Molecular Ecology*, *26*, 420–430.

Ambrose, S. H. (2001). Paleolithic technology and human evolution. *Science*, *291*, 1748–1753.

Avery, S. (2010). *Hydrological impacts of Ethiopias Omo basin on Kenya's Lake Turkana water levels and fisheries*. Final Report prepared for the African Development Bank, Tunis.

Barba, D., Miquel, C., Lobréaux, S., Quenette, P. Y., Swenson, J. E., & Taberlet, P. (2016). High-throughput microsatellite genotyping in ecology: Improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, *17*, 492–507.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (2004). *GENETIX 4.05, Logiciel Sous Windows TM Pour la Génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier.

Butzer, K. W. (1971). *Recent history of an Ethiopian delta – The Omo River and the level of Lake Rudolf*. Department of Geography, Research Paper No. 136. Chicago, USA: University of Chicago Press.

Carleton, K. L., Streelman, J. T., Lee, B. Y., Garnhart, N., Kidd, M., & Kocher, T. D. (2002). Rapid isolation of CA microsatellites from the tilapia genome. *Animal Genetics*, *33*(2), 140–144.

Clement, M., Posada, D., & Crandall, K. A. (2000). TSC: A computer program to estimate gene genealogies. *Molecular Ecology*, *9*, 1657–1659.

Coe, M. (1971). Notes on the *Tilapia nilotica* subsp. from the Kinyang (Ngingyan) and Suguta Rivers in the vicinity of the Kapedo alkaline hot springs, Kenya. *Revue de Zoologie et de Botanique Africaines*, *83*, 309–316.

Cornuet, J. M., & Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, *144*(4), 2001–2014.

Frankham, R. (1996). Relationship of genetic variation to population size in wildlife. *Conservation Biology*, *10*, 1500–1508.

Frostick, L. E., & Reid, I. (1990). Structural control of sedimentation patterns and implications for the economic potential of the east African rift basins. *Journal of African Earth Sciences*, *10*, 307–318.

Fuchs, V. E. (1950). Pleistocene events in the Baringo basin, Kenya Colony. *Geological Magazine, London*, *87*, 149–174.

Garcin, Y., Junginger, A., Melnick, D., Olago, D. O., Strecker, M. R., & Trauth, M. H. (2009). Late Pleistocene-Holocene rise and collapse of Lake Suguta, northern Kenya rift. *Quaternary Science Reviews*, *28*, 911–925.

Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, *10*, 305–318.

Gonzalez-Castellano, I., Gonzalez-Lopez, J., Gonzalez-Tizon, A. M., & Martinez-Lage, A. (2020). Genetic diversity and population structure of the rockpool shrimp *Palaemon elegans* based on microsatellite: Evidence for a cryptic species and differentiation across the Atlantic-Mediterranean transition. *Scientific Reports*, *10*, 10784.

Goudet, J. (1995a). FSTAT (vers. 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, *86*, 485–186.

Goudet, J. (1995b). *FSTAT: A program to estimate and test gene diversities and fixation indices version 2.9.3. 2001*. http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet.

Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, *27*(2), 221–224.

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., Cariani, A., Maes, G. E., Diopere, E., Carvalho, G. R., & Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources*, *11*, 123–136.

Heras, S., Planella, L., García-Marín, J. L., Vera, M., & Roldán, M. I. (2019). Genetic structure and population connectivity of the blue and red shrimp Aristeus antennatus. *Scientific Reports*, *9*, 1–11.

Hopson, A. J. (1982). *Lake Turkana: A report on the findings of the Lake Turkana project 1972–1975* (Vol. *16*). Overseas Development Administration.

Hundertmark, K. J., & Van Daele, L. J. (2010). Founder effect and bottleneck signatures in an introduced, insular population of elk. *Conservation Genetics*, *11*, 139–147.

Jorde, P. E., Sovic, G., Westgaard, J., Albretsen, J., André, C., Hvingel, C., Johansen, T., Sandvik, A. D., Kingsley, M., & Jørstad, K. E. (2015). Genetically distinct populations of northern shrimp, Pandalus borealis, in the North Atlantic: Adaptation to different temperatures as an isolation factor. *Molecular Ecology*, *24*, 1742–1757.

Kolding, J. (1989). *The fish resources of Lake Turkana and their environment. Report off KEN 043 trial fishery 1986–1987* (p. 262). University of Bergen. Report to NORAD, Oslo, December 1989.

Kolding, J. (1993). Population dynamics and life-history styles of Nile tilapia, *Oreochromis niloticus*, in Ferguson's gulf, Lake Turkana, Kenya. *Environmental Biology of Fishes*, *37*, 25–46.

Ledoux, J. B., Ghanem, R., Horaud, M., López-Sendino, P., Romero-Soriano, V., Antunes, A., Bensoussan, N., Gómez-Gras, D., Linares, C., Machordom, A., Ocaña, O., Templado, J., Leblois, R., Ben Souissi, J., & Garrabou, J. (2021). Gradient of genetic diversity and differentiation across the distribution range of a Mediterranean coral: Patterns, processes and conservation implications. *Diversity and Distributions*, *11*(27), 2104–2123.

Lee, W. J., & Kocher, T. D. (1996). Microsatellite DNA markers for genetic mapping in *Oreochromis niloticus*. *Journal of Fish Biology*, *49*, 169–171.

Leigh, J. W., & Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*, 1110–1116.

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*, 1451–1452.

Meyer, A., Kocher, T. D., Basasibwaki, P., & Wilson, A. C. (1990). Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature*, *347*(6293), 550–553.

Mworia, M. J. (1991). Vegetation response to climate change in central Rift Valley. *Quaternary Research*, *35*, 234–245.

Ndiwa, C. T. (2014). *Contribution to the knowledge of the populations of Nile tilapia (Oreochromis niloticus) inhabiting extreme conditions of temperature and alkalinity*. PhD Thesis, University of Montpellier II – Sciences et Techniques du Languedoc, 2Montpellier, France.

Ndiwa, C. T., Nyingi, D., Claude, J., & Agnèse, J. F. (2016). Morphological variations of wild populations of Nile tilapia found in extreme environmental conditions of Kenyan RiftValley. *Environmental Biology of Fishes*, *99*, 473–485.

Nyingi, D., De-Vos, L., Aman, R., & Agnèse, J. F. (2009). Genetic characterization of an unknown and endangered native population of the Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758) (Cichlidae; Teleostei) in the Loboi swamp (Kenya). *Aquaculture*, *297*, 57–63.

Owen, R. B., Barthelme, J. W., Renaut, R. W., & Vincens, A. (1982). Palaeolimnology and archeology of Holocene deposits north east of Lake Turkana, Kenya. *Nature*, *298*, 523–528.

Owen, R. B., Crossley, R., Johnson, T. C., Tweddle, D., Kornfield, I., Davison, S., Eccles, D. H., & Engstrom, D. E. (1990). Major lowlevels of Lake Malawi and their implication for speciation ratesin cichlid fishes. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *240*, 519–553.

Perina, A., Mari-Mena, N., Torrecilla, Z., Gonzalez-Tizon, A. M., González-Castellano, I., González-Ortegón, E., & Martínez-Lage, A. (2019). Assessment of genetic diversity and population structure of the common littoral shrimp Palaemon serratus (Pennant, 1777) by microsatellites: Towards a sustainable management. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *29*(4), 528–536.

Piry, S., Luikart, G., & Cornuet, J. M. (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, *90*(4), 502–503.

Sage, R. D., Loiselle, P. V., Basasibwaki, P., & Wilson, A. C. (1984). Molecular versus morphological change among cichlid fishes of Lake Victoria. In A. A. Echelle & I. Kornfield (Eds.), *Evolution of fish species flocks* (pp. 185–197). Univ. Maine Press.

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

Trewavas, E. (1983). *Tilapiine fishes of the genera Sarotherodon Oreochromis and Danakilia* (p. 583). British Museum (Natural History).

Umer, M., Kebede, S., & Osmaston, H. (2004). Quaternary glacial activity on the Ethiopian mountains. In J. Ehlers & P. L. Gibbard (Eds.), *Quaternary glaciations extent and chronology, part III* (pp. 171–174). Elsevier.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). MicroChecker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*, 535–538.

Yuretich, R. F. (1979). Modern sediments and sedimentary processes in Lake Rudolf (Lake Turkana) eastern rift valley, Kenya. *Sedimentology*, *26*, 313–331.

Yuretich, R. F., & Cerling, T. E. (1983). Hydrochemistry of Lake Turkana, Kenya-mass balance and mineral reactions in an alkaline lake. *Geochimica et Cosmochimica Acta*, *47*, 1099–1109.

How to cite this article: Ndiwa, T. C., Nyingi, D. W., & Agnèse, J.-F. (2023). Genetic diversity of a cichlid fish population after 100 years of isolation. *African Journal of Ecology*, *61*, 595–605. https://doi.org/10.1111/aje.13144