Journal of FISH BIOLOGY

fsbi An International Society for Fish Biology

Journal of Fish Biology (2016) **89**, 2557–2570
doi:10.1111/jfb.13150, available online at wileyonlinelibrary.com

High genetic diversity and population differentiation in *Clarias gariepinus* of Yala Swamp: evidence from mitochondrial DNA sequences[a]

J. E. Barasa*†, R. Abila‡, J. P. Grobler§, M. Agaba||¶, E. J. Chemoiwa** and B. Kaunda-Arara*

**Department of Fisheries and Aquatic Sciences, University of Eldoret, P. O. Box 1125-30100, Eldoret, Kenya, ‡Department of Environmental Studies, Forestry and Agriculture, Maasai Mara University, P. O. Box 861-20500, Narok, Kenya, §Department of Genetics, University of the Free State, P. O. Box 399, Bloemfontein 9300, South Africa, ||Department of Genetics and Genomics, Nelson Mandela African Institute of Science and Technology, P. O. Box 447, Arusha, Tanzania, ¶Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI Hub), P. O. Box 30709, Nairobi 00100, Kenya and **Department of Biological Sciences, University of Eldoret, P. O. Box 1125-30100, Eldoret, Kenya*

In order to improve the conservation and sustainable utilization of the African catfish *Clarias gariepinus* of the Yala Swamp in Kenya, genetic diversity and population structure of Lakes Kanyaboli and Namboyo populations of the species were studied using DNA sequences of the mitochondrial D-loop control region. Genetic diversity inferred as haplotype and nucleotide diversities and number of singletons and shared haplotypes was higher in the Lake Kanyaboli population (LKG) than the Lake Namboyo population (LNG) of *C. gariepinus*. Thirty-one haplotypes were inferred, of which 25 (80·6%) were private or singletons, while only six (19·4%) haplotypes were shared between LKG and LNG. Both populations were differentiated, with F_{ST} value that was significantly different from zero ($P < 0{\cdot}05$). Two clusters were inferred both from the maximum likelihood tree and the spanning networks of phylogenetic relationships of haplotypes. Mismatch distribution for total sample was multi-modal but individually, distributions were uni-modal in LKG, but multimodal in LNG. The mean ± s.d. raggedness index for both populations was 0·085 ± 0·098 and not significantly different from zero ($P > 0{\cdot}05$). Individual raggedness indices were 0·015 and 0·154 for LKG and LNG respectively. Fu's F_s was negative for both populations, with LKG recording −14·871, while LNG had −2·565, significantly different from zero for LKG ($P < 0{\cdot}05$), but the value for LNG was not significant ($P > 0{\cdot}05$). Tajima's D was negative for both populations, with LKG recording −1·734, while LNG had −1·136. Standardized square differences (SSD) were 0·001 for LKG and 0·048 for LNG and non-significant between them ($P > 0{\cdot}05$). Values between all populations were also not significantly different ($P > 0{\cdot}05$), mean ± s.d. SSD 0·025 ± 0·033.

Key words: African catfish; bait; haplotypes; mismatch distribution; mitochondrial D-loop; satellite lakes.

†Author to whom correspondence should be addressed. Tel.: +254 725 842098; email: jbarasa2002@yahoo.com

All authors made an equal contribution to this work.

[a]This paper was presented at the FSBI Symposium, Bangor, in July 2016. Its content may not follow the usual style and format of the *Journal of Fish Biology*.

INTRODUCTION

Genetic diversity is important for the persistence of a species in the environment (Lande, 1988). Loss of genetic diversity can be rapid in a small population, where genetic drift can erode genetic variation in a few generations (Frankharm *et al.*, 2002). In fish species, small populations may arise from the physical size of the aquatic habitat, anthropogenic activities such as overexploitation of the fish species, habitat destruction and fragmentation and the introduction of alien species. The relative abundance of a fish population in a fishery, however, is a function of catch per unit effort (Mwakubo *et al.*, 2007), which is often unknown in data deficient fisheries. These factors affect the demographic dynamics of a population (Laroche & Durand, 2004) and consequently the genetic diversity of the population. The consequences of these anthropogenic influences are often a reduction in genetic diversity of the fish population, which potentially affects the evolutionary ability and persistence of the population in the habitat.

A recent study reported higher genetic diversity in the Lake Victoria population of the African catfish *Clarias gariepinus* (Burchell 1822) than the Lake Kanyaboli population of the species (Barasa *et al.*, 2014). Lake Kanyaboli with a surface area of 10·5 km^2 is a satellite of and therefore a fragment of Lake Victoria (surface area of 69 000 km^2). Barasa *et al.* (2014) attributed the lower genetic diversity of Lake Kanyaboli population to a smaller population size. Therefore, satellite lakes are an example of aquatic habitats where size-related effects on genetic diversity of fish populations can be studied. Several such lakes abound in the Lake Victoria basin and in addition to Lake Kanyaboli include Lakes Namboyo and Sare in Kenya, Lake Nabugabo in Uganda and Lake Kirumi in Tanzania. Naturally small and isolated, satellite lakes have long been recognized as functional refugia for indigenous fish species (Kaufman & Ochumba, 1993; Kaufman *et al.*, 1997), whose survival in the main lake is severely threatened or may be extirpated altogether, due to anthropogenic effects. The loss of ichthyofaunal biodiversity in the Lake Victoria basin has called for an urgent need to identify and conserve the remaining populations, including the genetic diversity they contain, as well as their habitats (Kaufman *et al.*, 1997; Ojwang *et al.*, 2007; Chemoiwa *et al.*, 2013). In their study, Barasa *et al.* (2014) reported no significant differentiation between Lakes Victoria and Kanyaboli populations of *C. gariepinus*, but six private haplotypes were inferred in the Lake Kanyaboli population. The study attributed the presence of private haplotypes in Lake Kanyaboli to the possibility of the six haplotypes having been lost from Lake Victoria due to predation pressure from the exotic piscivore Nile perch *Lates niloticus* (L. 1758). Alternatively, the haplotypes could have arisen *in situ* in the Lake Kanyaboli population. This suggested the potential role of Lake Kanyaboli in the conservation of genetic resources of indigenous fish species of the Lake Victoria basin. Existence of unique haplotypes has also been reported for the Lake Kanyaboli populations of haplochromine cichlids (Abila *et al.*, 2004).

Lake Namboyo, another satellite of Lake Victoria that is found in the extensive Yala Swamp of Kenya, has a surface area of 2 km^2, separated from Lake Kanyaboli by a 6 km stretch of dense papyrus swamp and harbours *C. gariepinus*. Generally, *C. gariepinus* is an important predator in the ecosystem (Corbet, 1961) and in the satellite lakes of the Yala Swamp, the species is also exploited for human food, as bait for the *L. niloticus* long-line fishery in Lake Victoria and as brood stock for commercial aquaculture by the local community.

Fishing in the satellite lakes is the main activity of the local community (Aloo, 2003; Mwakubo *et al.*, 2007), with overfishing being reported in the two lakes (Aloo, 2003). Despite daily fishing expeditions in the satellite lakes, however, fisheries data are reported only for Lake Kanyaboli. In 2013 for instance, 38 t of *C. gariepinus* (19·5% of the total fishes landed) was landed from Lake Kanyaboli (State Department of Fisheries, 2013). In light of the sizes of the two satellite lakes, it is unclear how current rates of exploitation of *C. gariepinus* would affect genetic diversity of the species inhabiting both lakes. Similarly, since both lakes are fragments of Lake Victoria, with close geographical location and connectivity by the Yala Swamp, *C. gariepinus* populations of both lakes may not show significant population differentiation, because of a similar ancestor, younger evolutionary age having been separated from Lake Victoria 4000 years ago (Greenwood, 1965) and possibility of homogenization by gene flow due to dispersal of the fish.

Preservation of genetic diversity and identification of specific genetic units for management are important goals in biodiversity conservation programmes (Lesica & Allendorf, 1995). Genetic diversity of *Clarias* species would also help to identify suitable brood stock for improved aquaculture production. Aquaculture and propagation of *C. gariepinus* in hatcheries to supply seed for use as bait for *L. niloticus* in Lake Victoria has been suggested as a sustainable strategy to conserve the indigenous fish species of the Lake Victoria basin (Kaufman & Ochumba, 1993; Mkumbo & Mlaponi, 2007). It will also increase income of *C. gariepinus* farmers and hatchery operators (Chitamwebwa *et al.*, 2009; Barasa *et al.*, 2014). This study reports on the genetic diversity of *C. gariepinus* in Lakes Kanyaboli and Namboyo, satellites of Lake Victoria, Kenya, to test the effect of population size on genetic diversity and differentiation of the populations.

MATERIALS AND METHODS

Samples of *C. gariepinus* were collected from two satellite lakes of Lake Victoria in the Yala Swamp, 44 samples from Lake Kanyaboli (LKG) and 47 samples from Lake Namboyo (LNG) (Fig. 1). Comparable fish sample sizes have been used in related studies on the Asian walking catfish *Clarias batrachus* (L. 1758) of the Indonesian Archipelago (Pouyard *et al.*, 1998), the haplochromine cichids of Yala Swamp (Abila *et al.*, 2004) and the striped catfish *Pangasianodon hypophthalmus* (Sauvage 1878) of Thailand (Na-nakorn & Moeikum, 2009), although it would be more desirable to use higher sample sizes of fish. Fish were caught from both lakes using gillnets, traps and hook and line. Fin clips were obtained from fish samples and immediately preserved in 95% alcohol in clean cryovial tubes. A sequence of the Asian walking catfish *Clarias macrocephalus* Günther 1864 was obtained from GenBank (accession number FJ-495103) and used as an out-group for phylogenetic analysis.

DNA EXTRACTION

Genomic DNA was extracted from *c.* 25 mg of fin clip tissue using the Invitrogen PureLink (www.invitrogen.com) genomic DNA mini kit, according to the manufacturer's instructions. The purity and concentration of eluted DNA was determined by spectrophotometry using a Nanodrop 2000-Spectrophotometer (Thermo Scientific; www.thermoscientific.com). Extracted DNA was stored at −20° C until required for further analysis.

PCR AMPLIFICATION

DNA samples were amplified by PCR in a ABI 9700 thermal cycler (Thermo Scientific) using a pair of mitochondrial D-loop primers (forwards primer, L16473 5′-CTAAAAGCATCGG

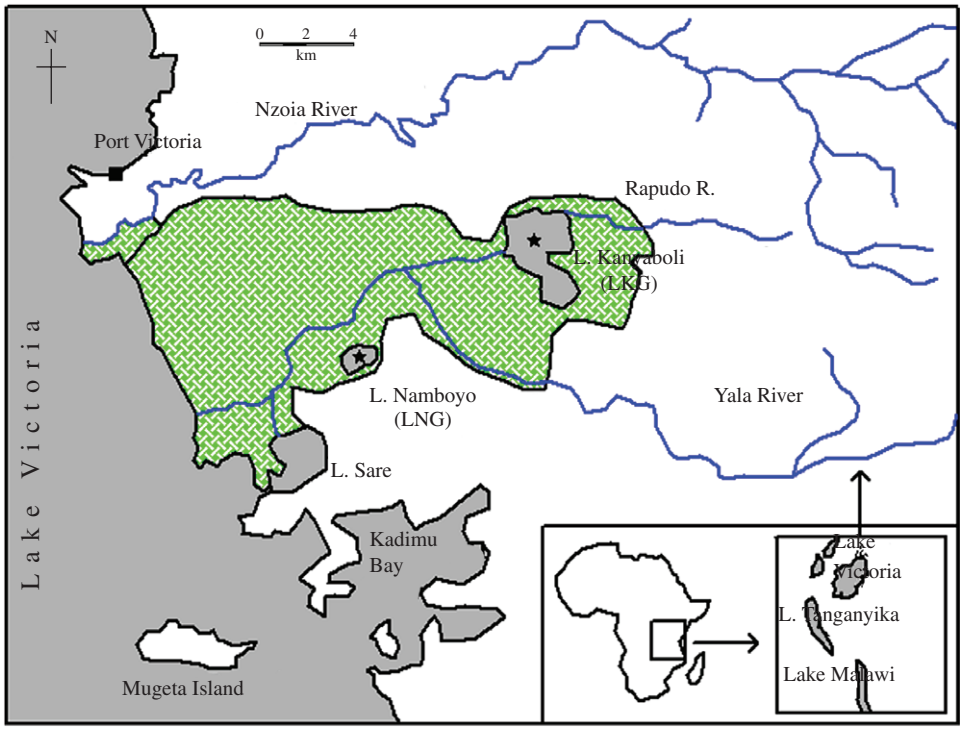

FIG. 1. Location of Lakes Kanyaboli and Namboyo in Yala Swamp of Lake Victoria, Kenya. Samples of *Clarias gariepinus* were taken from both Lakes Kanyaboli (LKG) and Namboyo (LNG). ★, Sampling sites. (Map adapted from Abila *et al*., 2008).

TCTTGTAATCC-3′; reverse primer, H355 5′-CCTGAAATGAGGAGGAACCAGATG-3′ (Nazia *et al*., 2010), targeting a product of 500 bp. PCR were done using AccuPower PCR PreMix (Bioneer; www.eng.bioneer.com), in 20 μl containing 10 μM each of forward and reverse primers and 50 ng DNA. PCR conditions were as described by Nazia *et al.* (2010) and used by Barasa *et al*. (2014, 2016). Annealing temperature was 56° C. The success of the PCR reaction was confirmed by 2·0% agarose gel electrophoresis of PCR products. PCR products were purified using the GeneJet PCR purification kit (Thermo Scientific), following the manufacturer's protocol. The quality and concentration of eluted and purified samples were checked by electrophoresis (2·0% agarose gel) and Nanodrop spectrophotometry (www.nanodrop.com).

The forward primer L16473 was used to sequence the PCR product of the D-loop region. The BigDye terminator premix sequencing kit (Thermo Scientific) was used for sequencing reactions, following the manufacturer's protocol. Products of the sequence reaction were cleaned by precipitation in absolute alcohol, re-suspended in Hi Di Formamide (Thermo Scientific) and BigDye terminator premix and run on the ABI 3730xl genetic analyser (www.thermofisher.com) with a capillary length of 50 cm.

DATA ANALYSIS

Data were analysed as previously described by Barasa *et al.* (2014), with DNA sequences aligned, assembled and trimmed in Bioedit 7.0.5 (Hall, 2005). For LKG a total of 44 sequences of the D-loop control region were used, comprising 16 sequences (Accession numbers KC594205–KC594220) from Barasa *et al.* (2014) and 28 new sequences from additional sampling in Lake Kanyaboli. LNG had 47 sequences. All sequences were trimmed to 401 bp. The

sequences were deposited in the GenBank database and are publicly available under accession numbers KU759960–KU759985 for LKG and KU759986–KU760032 for LNG. Duplicate haplotypes were identified using DNASP 5 (Librado & Rozas, 2009). Genetic diversity within populations was determined using DNASP and Arlequin 3.5 (Excoffier *et al.*, 2005), while genetic differentiation between groups was determined with Arlequin, expressed as F_{ST} values (Wright, 1965). The MEGA software 6.0.6 (Tamura *et al.*, 2007) was used to construct the maximum likelihood (ML) tree with 1000 bootstrap repeats and a D-loop sequence of *C. macrocephalus* (GenBank accession number FJ495103) was used as the out-group. Finally, the minimum spanning network showing the phylogenetic relationships among haplotypes was drawn using Network 4.56 (Bandelt *et al.*, 1999), with a median joining approach. Mismatch distributions exploiting pair wise differences among the D-loop control region sequences (Harpending, 1994) were used to investigate evidence of demographic changes among the samples, using DNASP 5.0. Uni-modal distributions indicate a recent population expansion, while more ragged distributions are indicative of a relatively stable population (Rogers & Harpending, 1992).

The population demographic expansion indices, Fu's F_s (Fu, 1997) and Tajima's D (Tajima, 1989), which are coalescent-based neutrality estimators, were generated using Arlequin 3.5. Two tests of goodness of fit, the standardized square difference (SSD) and the raggedness index were used to determine the fit between the observed and estimated distributions under a sudden population expansion model. The two tests determine departures from mutation-drift equilibrium, with negative values indicating an excess of new mutations due to either selective pressure or population expansion after a bottleneck. On the other hand, a positive value of the tests may indicate a stable population due to presence of an excess of old mutations.

RESULTS

The two populations had a total of 31 haplotypes, out of which 25 (80·6%) were private or singletons. Only 19·4% of the haplotypes (or six haplotypes) were shared, between LKG and LNG. These were haplotypes 1, 7, 11, 13, 14 and 22. Haplotype 1 was also the most abundant, with a total of 32 sequences (35·2% of total number of haplotypes), from both LKG and LNG. LKG had a higher number of singletons at 16 (51·6%) of the haplotypes, while LNG had nine (29·0%) singletons. Mean ± S.D. haplotype diversity was 0·888 ± 0·040 in LKG and 0·832 ± 0·043 in LNG, while the mean ± S.D. nucleotide diversity was 0·007 ± 0·002 and 0·010 ± 0·001 in the two populations respectively (Table I). The number of segregating sites was higher in LNG at 31, compared with LKG with 26 (Table I).

A total of two clusters of haplotypes were inferred from the ML tree (Fig. 2), with the two populations of LKG and LNG forming the clades, supported by bootstrap values of 40–60%. The two clades of LKG and LNG were also inferred in the minimum spanning network for haplotypes (Fig. 3), supporting the topology of the ML tree. In the network, haplotype 1 was the most abundant and all other haplotypes were connected to it with one to two mutation steps. The pair-wise comparisons for population differentiation index (F_{ST}) value for the two populations was 0·121 and was significantly different ($P < 0·01$) from zero.

From the mismatch distributions of pair-wise comparisons of nucleotides, the total sample for both populations showed a multi-modal distribution (Fig. 4). For individual populations, however, a uni-modal distribution was observed for LKG (Fig. 5), while a multi-modal distribution was recorded in LNG (Fig. 6). The raggedness index was 0·015 for LKG and 0·154 for LNG (Table I) and these values were not significantly different from zero for both populations ($P > 0·05$). The mean ± S.D. raggedness index for both samples, however, was 0·085 ± 0·098, which was not significantly different from

TABLE I. Values of *Clarias gariepinus* diversities from Lake Kanyaboli (LKG) and Lake Namboyo (LNG). Sample size is the number of sequences of each population used in the analysis. Standardized-square difference is a measure of goodness of fit between observed and expected distributions under a sudden population expansion model

	LKG	LNG
Sample size (n)	44	47
Mean ± S.D. haplotype diversity (h)	0·888 ± 0·040	0·832 ± 0·043
Nucleotide diversity (π)	0·007 ± 0·001	0·010 ± 0·001
Mean ± S.D. number of singletons	16	9
Number of shared haplotypes	6	6
Number of segregating sites	26	31
Raggedness index	$0{\cdot}015^{NS}$	0·154*
Fu's F_s	−14·871**	$-2{\cdot}565^{NS}$
Tajima's D	−1·734*	$-1{\cdot}136^{NS}$
Standardized-square difference	$0{\cdot}001^{NS}$	$0{\cdot}048^{NS}$

$^{NS}P > 0{\cdot}05$.
$^{*}P < 0{\cdot}05$.
$^{**}P < 0{\cdot}001$.

zero ($P > 0{\cdot}05$). The population demographic expansion index Fu's F was negative for both populations, with LKG recording −14·871, while LNG had −2·565. These values were significantly different from zero for LKG ($P < 0{\cdot}01$), but not significantly different from zero for LNG ($P > 0{\cdot}05$). On the other hand, the index Tajima's D was also negative for both populations, with LKG recording −1·734, while LNG had −1·136. Both indices were not significantly different from zero ($P > 0{\cdot}05$). The SSD, which is a goodness of fit test statistic between the expected and observed distributions, was 0·001 for LKG and 0·048 for LNG, none of which showed significant difference from zero ($P > 0{\cdot}05$). The mean ± S.D. SSD value for the two populations was 0·025 ± 0·033, which was also not significantly different from zero ($P > 0{\cdot}05$).

DISCUSSION

GENETIC DIVERSITY

The Lake Victoria basin has experienced one of the highest losses of species in the 21st century and the ichthyofaunal diversity of the basin is one of the most globally threatened. Conservation of the remaining habitats and populations is paramount in securing this source of food and livelihood for the local communities. This study assessed genetic diversity and population structure of *C. gariepinus* from two satellites of Lake Victoria, lakes with fish populations increasingly threatened by ongoing agri-based land-use changes. The conspecific populations of LKG and LNG had higher haplotype diversities (h) and low nucleotide diversities (π). This suggested that the *C. gariepinus* populations had undergone expansion after a bottleneck (Grant & Bowen, 1998; Avise, 2000) in which the effective population size declined. Rapid population expansion following a period of low effective population size is accompanied with new mutations (Avise *et al.*, 1984; Rogers & Harpending, 1992), reflected

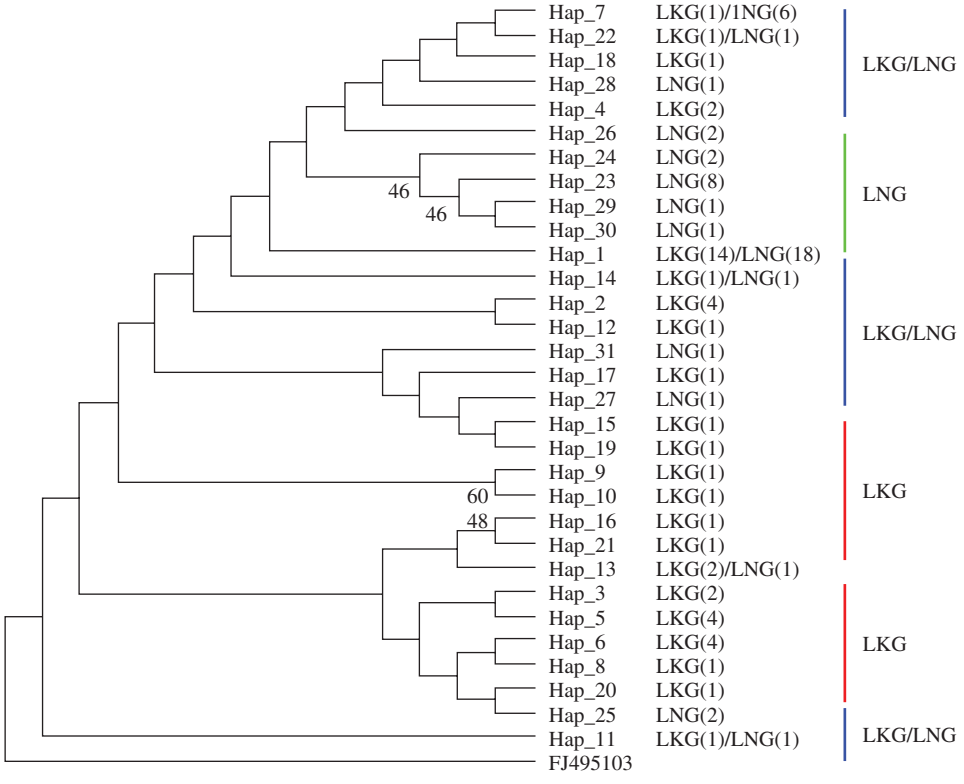

FIG. 2. Maximum likelihood tree for the Lake Kanyaboli population (LKG) Lake Namboyo population (LNG) *Clarias gariepinus*. Numbers in parentheses are frequency of haplotypes in LKG and LNG samples. Numbers on the nodes represent per cent boostrap values, based on 1000 bootstrap iterations. Bootstrap values below 40% were excluded. FJ495103 is the outgroup, a D-loop sequence of *Clarias macrocephalus*.

in the rise in haplotype diversities. The population bottleneck could have occurred during colonization of the satellite lakes, with only a few individuals after the formation of the lakes 4000 years ago (Greenwood, 1965). Rapid population expansion after colonization may have been facilitated by suitable environmental conditions in the satellites and absence of exotic species such as *L. niloticus* that could otherwise cause predation-pressure induced population declines, similar to what happened in the main Lake Victoria. Predation pressure caused the decline of *C. gariepinus* and the disappearance of *Clarias liocephalus* Boulenger 1898 from Lake Victoria (Goudswaard & Witte, 1997). The satellite lakes have a comparatively better water quality than the main Lake Victoria, which has experienced high levels of pollution, leading to eutrophication (Lung'ayia *et al.*, 2000). Eutrophication favours the proliferation of blue-green algae, which replaced diatoms as the dominant algal species in Lake Victoria (Lung'ayia *et al.*, 2000), unlike the satellite lakes in the Yala Swamp (Maithya, 1998). Population expansion in the studied populations is further demonstrated by star-like phylogenies (Fig. 3), in which the rare haplotypes are separated by one or two mutation steps from the central or common haplotype.

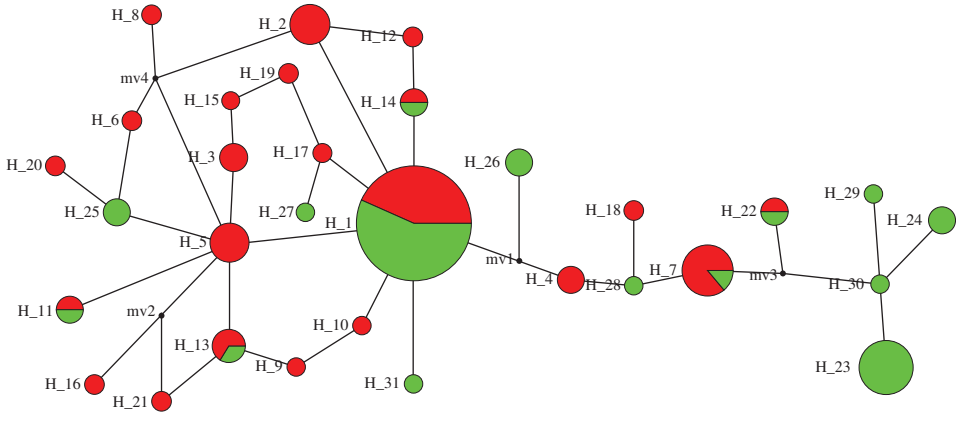

FIG. 3. Minimum spanning network showing the relationship between haplotypes of *Clarias gariepinus* from Lake Kanyaboli (LKG, ●), and from Lake Namboyo (LNG, ●) populations. Size of the circle is proportional to the haplotype frequency. Haplotypes are identified by the number of each haplotype. Branches represent the number of mutation steps.

LKG, however, had a higher genetic diversity than LNG, despite LNG having a slightly higher sample size (47) than LKG (44). This could be attributed to the fact that Lake Namboyo is much smaller (2 km^2) than Lake Kanyaboli (10·5 km^2), therefore *C. gariepinus* of Lake Kanyaboli has a higher population size than LNG. The study reported a high number of private haplotypes (singletons), at 80·6%, while only 19·4% were shared. This high number of singletons reflects high differentiation among populations, or limited gene flow. Limited gene flow among the two populations could be attributed to the presence of dense papyrus swamp fringing the lakes, which probably restricted the movement of the fish from one lake to the other. The restricted exchange of individuals isolated the lakes, which, coupled with the small size of the lakes, increased genetic drift in the populations, leading to differentiation between the populations. Small populations often suffer genetic drift (Frankharm *et al.*, 2002) and affect genetic differentiation of the populations. This differentiation of the two populations could also have been promoted by accumulation of mutations. The D-loop control region has a high rate of mutation and so accumulates changes in the genome (Meyer, 1994).

POPULATION GENETIC STRUCTURE

The study samples were differentiated or structured based on the sampling sites, supported by F_{ST} index that was significantly different from zero ($P < 0{\cdot}01$). This differentiation indicated high structuring among populations and limited gene flow. In the study of genetic diversity of *C. gariepinus* from Lakes Victoria and Kanyaboli, Barasa *et al.* (2014) reported similarity in the samples from the two sites, attributed to the fact that samples from Lake Kanyaboli were a fragment of Lake Victoria population. Anthropogenic-induced translocation of *C. gariepinus* juveniles from Lake Kanyaboli to Lake Victoria as bait for *L. niloticus* was also thought to contribute to this similarity. Contrary to the expected homogenization of the LKG and LNG populations of *C. gariepinus* on account of both being fragments of the Lake Victoria population,

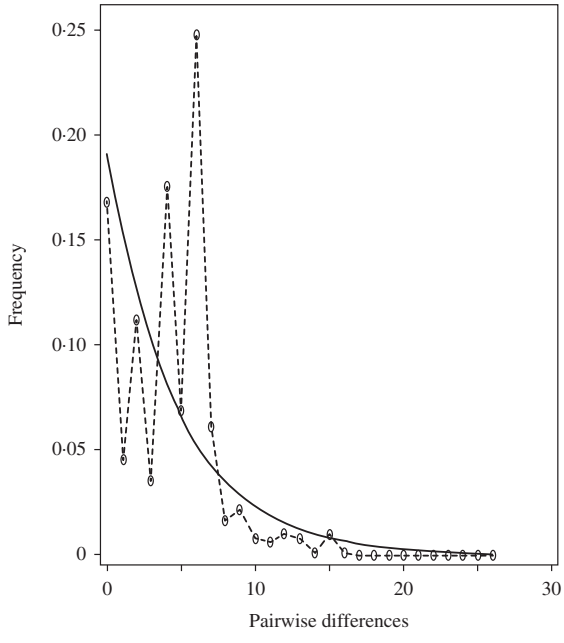

FIG. 4. Mismatch distributions from 401 bp sequences of the D-loop control region of the total *Clarias gariepinus* sample (both Lake Kanyaboli and Lake Namboyo population). ——, Expected distribution; ···◊···, observed distribution under a model of constant population size.

the F_{ST} index was significantly different from zero. First, limited gene flow reported in this study suggests the two populations were isolated.

Additionally, despite both populations being used as bait for *L. niloticus* in Lake Victoria, the translocation did not involve moving samples between the two lakes, rather directly between each lake and Lake Victoria. Although the trade in live bait in the Lake Victoria basin is poorly documented (Barasa *et al.*, 2014), bait traders collect *C. gariepinus* juveniles daily from the Yala Swamp lakes for sale to *L. niloticus* fishermen along Lake Victoria beaches. Similar genetic differentiation has been reported in three populations of the Nile tilapia *Oreochromis niloticus* (L. 1758) of the Loboi Swamp in the Lakes Baringo–Bogoria drainages in the Kenyan rift valley (Ndiwa *et al.*, 2014). The Lake Bogoria Hotel Spring, the Chelaba Spring and the Turtle Spring are separated only by a few hundred m to 1 km, but their respective *O. niloticus* populations reported significant F_{ST} indices and high genetic diversity, attributed to limited gene flow facilitated by the dense papyrus swamp fringing the springs and stable environmental conditions of the hot springs (Ndiwa *et al.*, 2014), in ways expected of populations isolated by distance.

DEMOGRAPHIC EXPANSION

For combined samples from both lakes, the mean raggedness index was higher than the mean SSD and since both values were not significantly different from zero, the fit of the expected and observed distributions to the model of population expansion was poor (Fig. 4). This discounted the null hypothesis of both populations having undergone

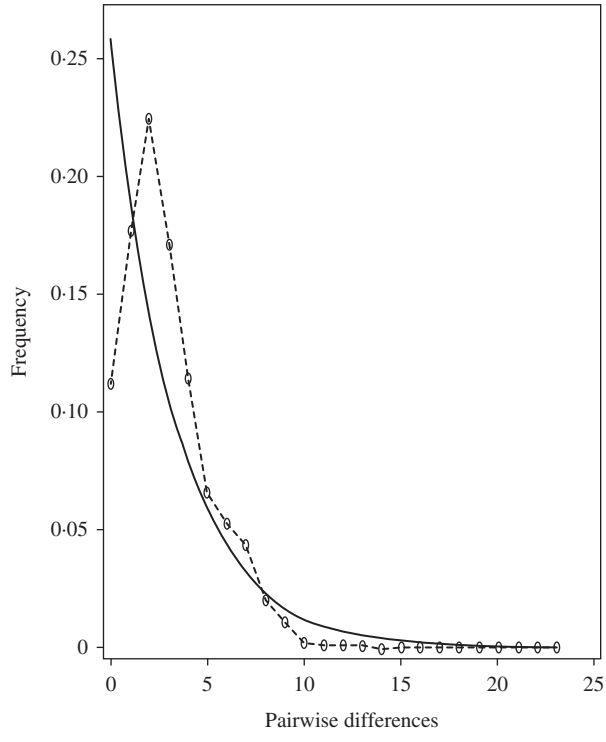

FIG. 5. Mismatch distribution from 401 bp sequences of the D-loop control region of Lake Kanyaboli *Clarias gariepinus* samples. ——, Expected distribution; ···◊···, observed distribution under a model of constant population size.

recent expansion. This trend was similar among the samples of individual populations, with the raggedness index being higher than SSD in LKG and LNG, indicating a poor fit of the expected and observed distributions to the population expansion model. The mismatch distribution of the total samples in the study was multi-modal and rough (Fig. 4). This showed that taken together, the samples under study are old and stable, without expansion (Rogers & Harpending, 1992). Multi-modal and more ragged distributions are a signature for more stable populations (Rogers & Harpending, 1992), reported in a number of fish taxa including *Labeobarbus* spp. Rüppell 1835 (Muwanika *et al.*, 2012), *Barbus altianalis* (Boulenger 1900) (Chemoiwa *et al.*, 2013) and *C. batrachus* (Lee & Sulaiman, 2015). LKG, however, showed a uni-modal and less ragged distribution, while LNG had a multi-modal and more ragged distribution, suggesting a recent population expansion in LKG, while LNG was a stable population.

Based on the population demographic expansion indices (Fu's F_s and Tajima's D), the Fu's F_s was negative for both populations, suggesting that the populations have probably undergone a recent expansion or are under purifying selection (Fu, 1997). Since both the raggedness index and SSD for LKG are not significantly different from zero ($P > 0{\cdot}05$), a purifying selection would be a more likely force influencing the demography of LKG, rather than a demographic expansion. A purifying selection increases mutations at a silent site, without drastically increasing the heterozygosity (Tajima, 1989). High exploitation pressure is reported in Lake Kanyaboli (Aloo,

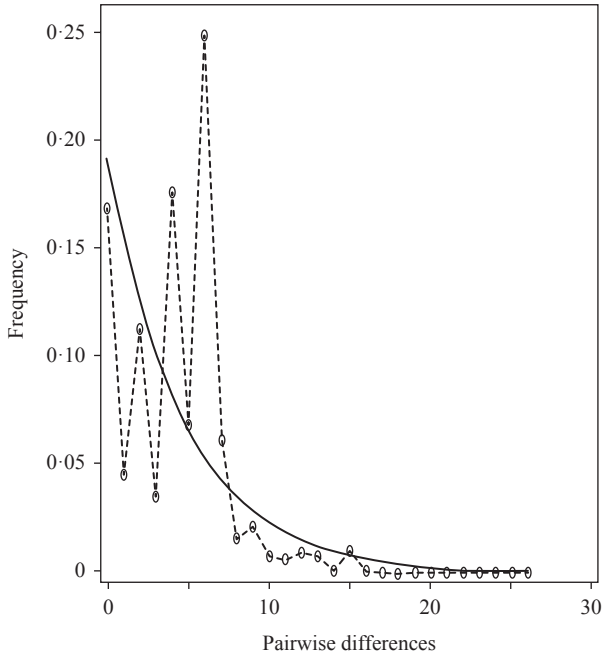

FIG. 6. Mismatch distributions of 401 bp of D-loop control region of Lake Namboyo population *Clarias gariepinus* samples. ——, Expected distribution; ···0···, observed distribution under a model of constant population size.

2003) to meet the high demand for food fishes, which for clariids in the lake has been exacerbated by the trade in *C. gariepinus* bait for *L. niloticus* in Lake Victoria (Barasa *et al.*, 2014). High exploitation pressure of natural fish populations lowers the size at first maturity, to increase reproduction, in order to compensate for the numbers of fishes lost by fishing mortality (Pukk *et al.*, 2013; Lappalainen *et al.*, 2016). This is supported by a negative *D* for all the populations, suggesting an increasing population size or a purifying selection at a locus (Tajima, 1989). The purifying selection that increases mutations at the locus could also be increasing genetic diversity at the locus, leading to the higher haplotype and nucleotide diversities in this study, compared with values reported by earlier studies in similar population (Barasa *et al.*, 2014) and different populations (Roodt-Wilding *et al*., 2010; Barasa *et al*., 2014, 2016).

This study has shown higher genetic diversity in the Lake Kanyaboli than Lake Namboyo C. *gariepinus* populations. Both satellite lakes should be conserved by reducing human activities that affect the ecological integrity of the Yala Swamp. Similarly, translocation of *C. gariepinus* between both lakes that could homogenize the populations should be avoided. The findings of the current study are relevant to the management of indigenous fish resources and the choice of *C. gariepinus* brood stock for commercial aquaculture in Kenya. While the satellite lakes studied are smaller with lower fish output compared with the major inland fisheries such as Lakes Victoria, Turkana and Baringo, they are connected to larger lakes through extensive swamps. Where human activity alters environmental conditions of the swamp, increased gene flow could homogenise previously differentiated populations of fish species. Similarly,

translocation of fish populations across drainage basins for aquaculture, bait and ornamental purposes could pollute unique gene pools. For instance, the Loboi Swamp harbours unique and previously unidentified populations of *O. niloticus baringoensis* Trewavas 1983 (Ndiwa *et al.*, 2014), which is seriously threatened by aquaculture in the Kapkuikui area of the basin and has *C. gariepinus* as well. Recently, many farmers in Kapkuikui area of the swamp and the upper parts of the Lake Baringo basin have been assisted by Baringo county fisheries authorities to stock ponds with *O. niloticus* and swamps and reservoirs with *C. gariepinus* seeds acquired from Dominion Farms limited (www.dominion-farms.com; J. E. Barasa, pers. obs.), one of the commercial fish farms in Yala Swamp area of Lake Victoria basin. Since Loboi Swamp is connected to Lake Baringo *via* the Molo River and with extensive irrigation canals of the Perkerra irrigation scheme, such a stocking exercise will compromise the genetic purity of Lake Baringo *C. gariepinus*, which is probably distinct from the Lake Victoria and Kanyaboli populations of the species (J. E. Barasa, S. Mdyogolo, R. Abila, P. J. Grobler, R. A. Skilton, M. N. Njahira, E. J. Chemoiwa, O. G. Dangasuk, B. Kaunda-Arara & E. Verheyen, unpubl. data).

This study was supported by the National Commission for Science and Technology (NACOSTI), Kenya, research grant to R.A. through the Kenya-South Africa Scientific and Research cooperation (SAK/2nd Call) and a Science, Technology and Innovation PhD research grant (NCST/5/003/3rd Call) to J.E.B. Genetic analysis was carried out at the biosciences eastern and central Africa-International Livestock Research Institute Hub (BecA-ILRI Hub) in Nairobi, Kenya. We are grateful for financial support provided to BecA-ILRI Hub by the Australian Department for Foreign Affairs and Trade (DFAT) through the BecA-CSIRO partnership, Syngenta Foundation for Sustainable Agriculture (SFSA), the U.K. Department for International Development (DFID), the Bill and Melinda Gates Foundation (BMGF) and the Swedish International Development Agency (SIDA). We also acknowledge the excellent technical support provided by Segolip, the BecA-ILRI Hub sequencing and genotyping research support unit. We thank M. N. Njahira for technical support during laboratory work.

References

Abila, R., Barluenga, M., Engelken, J., Meyer, A. & Salzburger, W. (2004). Population structure and genetic diversity in a haplochromine cichlid of a satellite lake of Lake Victoria. *Molecular Ecology* **13,** 2589–2602. doi: 10.1111/j.1365-294X.2004.02270.x

Abila, R., Salzburger, W., Ndonga, M. F., Owiti, D. O., Barluenga, M. & Meyer, A. (2008). The role of the Yala Swamp lakes in the conservation of the Lake Victoria region haplochromine cichlids: evidence from genetic and trophic ecology studies. *Lakes & Reservoirs: Research and Management* **13,** 95–104. doi: 10.1111/j.1440-1770.2008.00366.x

Aloo, P. A. (2003). Biological diversity of Yala Swamp lakes, with special emphasis on fish species composition, in relation to changes in the Lake Victoria Basin (Kenya): threats and conservation measures. *Biodiversity & Conservation* **12,** 905–920.

Avise, J. C. (2000). Phylogeography*:* The History and Formation of Species, 2nd edn. Cambridge, MA: Harvard University Press.

Avise, J. C., Neigel, J. E. & Arnold, J. (1984). Demographic influences on the mitochondrial DNA lineage survivorship in animal populations. *Journal of Molecular Evolution* **20,** 99–105.

Bandelt, H. J., Forster, P. & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16,** 37–48.

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. doi: 10.2989/16085914.2014.933734

Barasa *et al.* (2016). Genetic diversity and population structure in African catfish, Clarias gariepinus in Kenya: implication for aquaculture and conservation. *Begian Journal of Zoology*. (in press).

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.

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.

Frankharm, R., Ballou, J. D. & Briscooe, D. A. (2002). *Introduction to Conservation Genetics*. Cambridge: Cambridge University Press.

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

Goudswaard, K. P. C. & Witte, F. (1997). The catfish fauna of Lake Victoria after the Nile perch upsurge. *Environmental Biology of Fishes* **49,** 21–43.

Grant, W. S. & Bowen, B. W. (1998). Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of Heredity* **89,** 415–426.

Greenwood, P. H. (1965). The cichlid fishes of Lake Nabugabo Uganda. *Bulletin of the British Museum of Natural History* **12,** 313–357.

Hall, T. A. (2005). Bioedit: a user friendly biological sequence alignment editor and analysis software for windows 9598/NT. *Nucleic Acids Symposium Series* **41,** 95–98.

Harpending, H. C. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66,** 591–600.

Kaufman, L. & Ochumba, P. (1993). Evolutionary and conservation biology of cichlid fishes as revealed by faunal remnants in the northern Lake Victoria. *Conservation Biology* **7,** 719–730.

Kaufman, L. S., Chapman, L. J. & Chapman, C. A. (1997). Evolution in fast-forward: haplochromine fishes of the Lake Victoria region. *Endeavour* **21,** 23–29.

Lande, R. (1988). Genetics and demography in biological conservation. *Science* **241,** 1455–1460.

Lappalainen, A., Saks, L., Sustar, M., Heikinheimo, O., Jurgens, K., Kokkonen, E., Kurkilahti, M., Verliin, A. & Vetemaa, M. (2016). Length at maturity as a potential indicator of fishing pressure. *Fisheries Research* **174,** 47–57. doi: 10.1016/j.fishres.2015.08.013

Laroche, J. & Durand, J. D. (2004). Genetic structure of fragmented populations of a threatened endemic percid of the Rhone River: *Zingel asper*. *Heredity* **92,** 329–334.

Lee, P. & Sulaiman, Z. (2015). Genetic identification and structure of *Clarias batrachus* (Linnaeus 1758) from southern Asia using a mitochondrial DNA marker. *Zootaxa* **3962,** 182–190. doi: 10.11646/zootaxa.3962.1.11

Lesica, P. & Allendorf, F. W. (1995). When are peripheral populations viable for conservation? *Conservation Biology* **9,** 753–760.

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

Lung'ayia, H. B. O., M'Harzi, A., Tackx, M., Gichuki, J. & Symoens, J. J. (2000). Phytoplankton community structure and environment in the Kenyan waters of Lake Victoria. *Freshwater Biology* **43,** 529–543.

Maithya, J. (1998). A survey of the ichthyofauna of Lake Kanyaboli and other small water bodies in Kenya: alternative refugia for endangered fish species. *Naga, the ICLARM Quarterly* **1,** 54–56.

Meyer, A. (1994). DNA technology and phylogeny of fish. In *Molecular Systematics and Evolution of Marine Organisms* (Beumont, A. R., ed), pp. 219–248. London: Chapman & Hall.

Mkumbo, O. C. & Mlaponi, E. (2007). Impact of the baited hook fishery on the recovering endemic fish species in Lake Victoria. *Aquatic Ecosystem Health and Management* **10,** 458–466.

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

Mwakubo, S. M., Ikiara, M. M. & Abila, R. (2007). Socio-economic and ecological determinants in wetland fisheries in the Yala Swamp. *Wetlands Ecology and Management* **15,** 521–528.

Na-nakorn, U. & Moeikum, T. (2009). Genetic diversity of domesticated striped catfish, *Pangasianodon hypophthalmus* (Sauvage 1878), in Thailand: relevance to brood stock management regimes. *Aquaculture* **297,** 70–77. doi: 10.1016/j.aquaculture.2009.09.014

Nazia, A. K., Suzana, M., Azhar, H., Nguyen Thuy, T. T. & Siti Azizah, M. N. (2010). No genetic differentiation between geographically isolated populations of *Clarias macrocephalus* in Malaysia revealed by sequences of mtDNA cytochrome *b* and D-loop gene regions. *Journal of Applied Ichthyology* **26,** 568–570. doi: 10.1111/j.1439-0426.2010.01469.x

Ndiwa, T. C., Nyingi, D. W. & Agnese, J. F. (2014). An important natural resource of *Oreochromis niloticus* (Linnaeus 1758) threatened by aquaculture activities in the Loboi Drainage, Kenya. *PLoS ONE* **9,** e106972. doi: 10.1371/journal.pone.0106972

Ojwang, W. O., Kaufman, L., Soule, E. & Asila, A. A. (2007). Evidence of stenotopy and anthropogenic influence on carbon source for two major riverine fishes of the Lake Victoria watershed. *Journal of Fish Biology* **70,** 1430–1446.

Pukk, L., Kuparinen, A., Jarv, L., Gross, R. & Vasemagi, A. (2013). Genetic and life-history changes associated with fisheries-induced population collapse. *Evolutionary Applications* **6,** 749–760. doi: 10.1111/eva.12060

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

Roodt-Wilding, R., Swart, B. L. & Impson, N. D. (2010). Genetically distinct Dutch domesticated *Clarias gariepinus* used in aquaculture in southern Africa. *African Journal of Aquatic Science* **35,** 241–249. doi: 10.2989/16085914.2010.538507

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

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24,** 1596–1599.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19,** 395–420.

Electronic References

Excoffier, L., Laval, G. & Schneider, S. (2005). Arlequin: an integrated software package for population genetics data analysis. Version 3.0. *Evolutionary Bioinformatics Online* **1,** 47–50. Available at http://www.la-press.com/arlequin-version-30-an-integrated-software-package-for-population-gene-article-a188-abstract?article_id=188&tab=abstract/.

Pouyard, L., Hadie, W. & Surdato, E. (1998). Genetic diversity among *Clarias batrachus* (Siluriformes, Clariidae) populations from the Indonesian Archipelago. In *Proceedings of the Mid-Term Workshop of the Asia Catfish Project* (Legendre, M. & Pariselle, A., eds), pp. 43–48. Toulouse: ParaGraphic. Available at http://horizon.documentation.ird.fr/exl-doc/pleins_textes/doc34-08/010020339.pdf/.

State Department of Fisheries (2013). *Fisheries Annual Statistical Bulletin 2013*. Nairobi: Ministry of Agriculture, Livestock and Fisheries. Available at http://www.kilimo.go.ke/fisheries/wp-content/uploads/2015/05/Annual-Fisheries-Statistical-Bulletin-2013.pdf/.